



INFLUENCE DE LA TÂCHE SUR LE TRAITEMENT DES SCÈNES VISUELLES COMPLEXES : UNE MODÉLISATION DES NIVEAUX DE REPRÉSENTATION

Emmanuelle Boloix

► To cite this version:

Emmanuelle Boloix. INFLUENCE DE LA TÂCHE SUR LE TRAITEMENT DES SCÈNES VISUELLES COMPLEXES : UNE MODÉLISATION DES NIVEAUX DE REPRÉSENTATION. domain_other. Université de Provence - Aix-Marseille I, 2005. Français. NNT : . tel-00011364

HAL Id: tel-00011364

<https://theses.hal.science/tel-00011364>

Submitted on 12 Jan 2006

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

UNIVERSITÉ DE PROVENCE
29, AV. ROBERT SCHUMAN, 13621 AIX EN PROVENCE CEDEX 1
UFR PSYCHOLOGIE, SCIENCES DE L'ÉDUCATION

**INFLUENCE DE LA TÂCHE SUR LE TRAITEMENT DES SCÈNES
VISUELLES COMPLEXES :
UNE MODÉLISATION DES NIVEAUX DE REPRÉSENTATION**

Thèse pour obtenir le grade de
DOCTEUR DE L'UNIVERSITÉ DE PROVENCE
Formation doctorale : Psychologie

Présentée et soutenue publiquement par

Emmanuelle BOLOIX

Sous la direction de Claude BASTIEN

JURY

Claude BASTIEN, Professeur émérite, Université de Provence (Directeur)
Muriel BOUCART, Directeur de Recherche, CHRU Lille (Rapporteur)
Christian MARENDAZ, Professeur, Université Pierre Mendès France
Jean-François RICHARD, Professeur émérite, Université Paris 8
Charles TIJUS, Professeur, Université Paris 8 (Rapporteur)

DÉCEMBRE 2005

TABLE DES MATIÈRES

INTRODUCTION.....	- 9 -
-------------------	-------

PREMIÈRE PARTIE : ASPECTS THÉORIQUES

– CHAPITRE UN – LA PERCEPTION DE SCÈNES VISUELLES COMPLEXES.....	- 14 -
--	--------

1. IDENTIFICATION DE SCÈNES VISUELLES	- 15 -
1.1. <i>Résumé et arrangement spatial de la scène</i>	- 15 -
1.2. <i>Le schéma de scène</i>	- 17 -
2. ATTENTION ET PERCEPTION DES SCÈNES VISUELLES.....	- 20 -
2.1. <i>L'attention visuelle s'oriente vers certains stimuli visuels de manière exogène</i>	- 21 -
2.2. <i>La perception des scènes dépend de l'orientation endogène de l'attention visuelle</i>	- 28 -
3. EN RÉSUMÉ.....	- 36 -

– CHAPITRE DEUX – LES MODÈLES DE L'ORIENTATION ATTENTIONNELLE DANS LA PERCEPTION DE SCÈNES VISUELLES	- 38 -
---	--------

1. LES MODÈLES DE RÉFÉRENCE	- 40 -
1.1. <i>Le modèle théorique de l'intégration de dimensions par l'attention</i>	- 40 -
1.2. <i>Le modèle « Guided Search 2 »</i>	- 42 -
1.3. <i>Avantages et lacunes de ces modèles</i>	- 45 -
2. LES MODÈLES DE LA PERCEPTION DE SCÈNES NATURELLES	- 46 -
2.1. <i>Architecture triadique de la vision</i>	- 47 -
2.2. <i>Un modèle computationnel de l'attention visuelle</i>	- 50 -
2.3. <i>Avantages et inconvénients des modèles présentés</i>	- 53 -
3. ATTENTION... REPRÉSENTATION ?.....	- 56 -

– CHAPITRE TROIS – LA REPRÉSENTATION EN MÉMOIRE DES SCÈNES VISUELLES**COMPLEXES - 58 -**

1. LES FACTEURS INFLUENÇANT LA REPRÉSENTATION EN MÉMOIRE DES OBJETS VISUELS - 62 -
 - 1.1. *Intérêt sémantique de l'information* - 63 -
 - 1.2. *Pertinence de l'information visuelle par rapport à la tâche* - 66 -
2. LES REPRÉSENTATIONS VISUELLES : ÉPARSES OU DÉTAILLÉES ? - 72 -
 - 2.1. *Beaucoup d'appelées, peu d'élues : rares sont les informations visuelles à être représentées en mémoire* - 73 -
 - 2.2. *Des représentations visuelles détaillées, qui se rapprochent de la réalité visuelle* - 76 -
3. UNE HYPOTHÈSE ALTERNATIVE : PLUSIEURS NIVEAUX DE REPRÉSENTATION DE L'INFORMATION VISUELLE. - 79 -

– CHAPITRE QUATRE – PROBLÉMATIQUE - 83 -**DEUXIÈME PARTIE: ÉTUDES COMPORTEMENTALES****– CHAPITRE CINQ – EFFET DES FACTEURS SÉMANTIQUES ET FONCTIONNELS SUR LA****PERCEPTION DE SCÈNES VISUELLES COMPLEXES - 91 -**

1. EXPÉRIENCE 1 : EFFET DES FACTEURS FONCTIONNELS SUR LA PERCEPTION DE SCÈNES VISUELLES COMPLEXES - 93 -
 - 1.1. *Méthode* - 94 -
 - 1.1.1. Participants - 94 -
 - 1.1.2. Matériel - 94 -
 - 1.1.3. Équipement - 97 -
 - 1.1.4. Procédure : tâche de détection de changements - 97 -
 - 1.2. *Analyse des résultats* - 100 -

1.2.1.	Analyse des scores de détection.....	- 102 -
1.2.2.	Analyse des latences de réponses	- 103 -
1.2.3.	Analyse des temps d'exploration.....	- 104 -
1.3.	<i>Discussion</i>	- 104 -
2.	EXPÉRIENCE 2 : EFFET DES FACTEURS SÉMANTIQUES ET FONCTIONNELS SUR LA PERCEPTION DE SCÈNES NATURELLES COMPLEXES	- 109 -
2.1.	<i>Méthode</i>	- 111 -
2.1.1.	Participants	- 111 -
2.1.2.	Matériel	- 111 -
2.1.3.	Équipement.....	- 113 -
2.1.4.	Procédure : tâche de détection de changements.....	- 113 -
2.2.	<i>Résultats</i>	- 114 -
2.2.1.	Analyse des performances de détection au test de détection explicite	- 115 -
2.2.2.	Analyse des performances au test de détection implicite.....	- 121 -
2.3.	<i>Discussion</i>	- 124 -
3.	DISCUSSION DU CHAPITRE CINQ.....	- 131 -
 – CHAPITRE SIX – EFFET DE LA RÉOLUTION DE LA TÂCHE SUR LA REPRÉSENTATION DE L'INFORMATION VISUELLE EN MÉMOIRE		- 139 -
1.	EXPÉRIENCE 3 : INFLUENCE DE LA RÉOLUTION DE LA TÂCHE SUR LA DYNAMIQUE DES REPRÉSENTATIONS VISUELLES.....	- 141 -
1.1.	<i>Méthode</i>	- 142 -
1.1.1.	Participants	- 142 -
1.1.2.	Matériel	- 142 -
1.1.3.	Équipement.....	- 150 -
1.1.4.	Procédure : tâche de résolution de problème couplée à une tâche de détection des changements	- 150 -
1.2.	<i>Analyse des résultats</i>	- 154 -
1.2.1.	Situations « un changement »	- 154 -
1.2.2.	Situation « six changements simultanés ».....	- 163 -
1.3.	<i>Discussion</i>	- 166 -

2.	EXPÉRIENCE 4 : EFFET DU RANG DU SOUS-BUT MODIFIÉ SUR LA DÉTECTION DES CHANGEMENTS : UN	
	EFFET DE L'EXPLORATION VISUELLE (PARTIE 1) ?	- 172 -
2.1.	<i>Méthode</i>	- 173 -
2.1.1.	Participants	- 173 -
2.1.2.	Matériel	- 173 -
2.1.3.	Équipement.....	- 173 -
2.1.4.	Procédure : tâche de résolution de problème couplée à une tâche de détection des changements	- 174 -
2.2.	<i>Analyse des résultats</i>	- 175 -
2.3.	<i>Discussion</i>	- 175 -
3.	EXPÉRIENCE 5 : EFFET DU RANG DU SOUS-BUT MODIFIÉ SUR LA DÉTECTION DES CHANGEMENTS : UN	
	EFFET DE L'EXPLORATION VISUELLE (PARTIE 2) ?	- 177 -
3.1.	<i>Méthode</i>	- 177 -
3.1.1.	Participants	- 177 -
3.1.2.	Matériel	- 177 -
3.1.3.	Équipement.....	- 178 -
3.1.4.	Procédure : tâche de résolution de problème couplée à une tâche de détection des changements	- 178 -
3.2.	<i>Analyse des résultats</i>	- 179 -
3.3.	<i>Discussion</i>	- 180 -
4.	DISCUSSION DU CHAPITRE SIX	- 181 -

– CHAPITRE SEPT – MODÉLISATION DU RÔLE DES BUTS DE LA TÂCHE SUR LA REPRÉSENTATION DE L'INFORMATION VISUELLE EN MÉMOIRE..... - 184 -

1.	UN MODÈLE MULTINOMIAL DE LA REPRÉSENTATION DE SCÈNES VISUELLES COMPLEXES EN CONDITION DE VISION ACTIVE	- 186 -
1.1.	<i>Le modèle</i>	- 187 -
1.2.	<i>Ré-analyse des données des expériences 4 et 5</i>	- 191 -
1.2.1.	Aspects pratiques.....	- 191 -
1.2.2.	Résultats	- 193 -
1.3.	<i>Discussion</i>	- 198 -
2.	HYPOTHÈSES SUR L'INFLUENCE DE LA TÂCHE SUR LA REPRÉSENTATION DES OBJETS VISUELS ET DE LEURS PROPRIÉTÉS CONSTITUTIVES	- 200 -

3.	EXPÉRIENCE 6 : INFLUENCE DU POINT DE VUE ADOPTÉ POUR RÉSOUDRE LA TÂCHE SUR LA REPRÉSENTATION DES OBJETS VISUELS ET DE LEURS PROPRIÉTÉS CONSTITUTIVES	- 203 -
3.1.	<i>Méthode</i>	- 203 -
3.1.1.	Participants	- 203 -
3.1.2.	Matériel	- 204 -
3.1.3.	Équipement.....	- 208 -
3.1.4.	Procédure.....	- 208 -
4.	MODÉLISATION DE L'INFLUENCE DU POINT DE VUE SUR LA REPRÉSENTATION DES SCÈNES VISUELLES -	210
4.1.	<i>Les prédictions de notre modèle multinomial</i>	- 210 -
4.2.	<i>Analyse des résultats</i>	- 214 -
4.3.	<i>Discussion</i>	- 221 -
5.	DISCUSSION DU CHAPITRE SEPT	- 224 -

TROISIÈME PARTIE: DISCUSSION GÉNÉRALE

DISCUSSION GÉNÉRALE.....	- 228 -
1. SYNTHÈSE ET INTERPRÉTATION DES RÉSULTATS OBTENUS	- 230 -
2. UN MODÈLE THÉORIQUE DES NIVEAUX DE REPRÉSENTATION DES SCÈNES VISUELLES COMPLEXES....	- 233 -
3. PERSPECTIVES DE RECHERCHE.....	- 236 -
BIBLIOGRAPHIE.....	- 240 -
ANNEXES	- 257 -
ANNEXE 1 : CONSTRUCTION DU MATÉRIEL DE L'EXPÉRIENCE 1	- 258 -
ANNEXE 2 : EXEMPLE DE CHAQUE COMBINAISON DE CHANGEMENT (EXPÉRIENCE 2)	- 259 -
ANNEXE 3A : CARTES UTILISÉES (EXPÉRIENCES 3, 4, 5, 6)	- 261 -

ANNEXE 3B : MASQUE UTILISÉE (EXPÉRIENCES 3, 4, 5, 6).....	- 264 -
ANNEXE 4 : EXEMPLE DE CHANGEMENT GLOBAL, DANS LA CONDITION « DISPARITION » (EXPÉRIENCE 3).	- 265 -
ANNEXE 5 : NOMBRE DE RÉPONSES EN FONCTION DU RANG DU CHANGEMENT ET DES EXPÉRIENCES IMPLÉMENTÉ DANS LE MODÈLE MPT (EXPÉRIENCES 4 & 5)	- 266 -
ANNEXE 6 : ÉQUATIONS GÉNÉRÉES PAR LE MODÈLE MPT (EXPÉRIENCES 4 & 5)	- 267 -
ANNEXE 7 : ARBRES DES TÂCHES UTILISÉS DANS L'EXPÉRIENCE 6	- 269 -
ANNEXE 8 : PLAN EN CARRÉ LATIN DE L'EXPÉRIENCE 6.....	- 270 -
ANNEXE 9 : NOMBRE DE RÉPONSES EN FONCTION DU RANG DU CHANGEMENT ET DE LA COHÉRENCE POINT DE VUE / CHANGEMENT IMPLÉMENTÉ DANS LE MODÈLE MPT (EXPÉRIENCE 6)	- 271 -
ANNEXE 10 : ÉQUATIONS GÉNÉRÉES PAR LE MODÈLE MPT (EXPÉRIENCE 6)	- 272 -
ANNEXE 11 : UN MODÈLE MULTINOMIAL DES NIVEAUX DE REPRÉSENTATION DES SCÈNES VISUELLES COMPLEXES	- 274 -

Introduction

Le monde visuel qui nous entoure et dans lequel se développe chacune de nos actions est particulièrement riche et complexe. L'être humain y évolue avec aisance alors qu'il est impossible, même pour le meilleur des systèmes de traitement de l'information, de considérer précisément la profusion des éléments qui composent notre monde et les relations qu'ils entretiennent les uns les autres. Or, en dépit de limitations mnésiques et attentionnelles considérables, notre système cognitif gère ce flux continu d'informations avec une redoutable efficacité et nous offre l'impression d'un monde stable et cohérent (Blackmore, Brelstaff, Nelson, & Trocianko, 1995; Dennet, 1993; Levin, Drivdahl, Momen, & Beck, 2002; Levin, Momen, & Drivdahl, 2000; Noë, Pessoa, & Thompson, 2000; O'Regan & Noë, 2001). Bien entendu, il ne fait plus de doute aujourd'hui que la représentation que nous avons de notre monde visuel est à mille lieues d'un décalque de la réalité, qui serait examiné par un observateur central, lequel prendrait ensuite des décisions sur la réalité de ce que nous voyons (Dennet, 1993), au contraire¹ ! Pourtant, nul ne pourrait raisonnablement nier que nous sommes néanmoins experts dans le traitement des informations visuelles. A titre d'exemple, considérons le cas d'un automobiliste, qui au volant de son véhicule, doit rendre visite à des amis qui viennent d'emménager dans une ville qu'il n'avait jamais eu l'occasion de visiter.

¹ Certains auteurs vont même jusqu'à utiliser les termes de « grande illusion » pour désigner l'écart qui existe entre la richesse du monde qui nous entoure et la pauvreté des représentations visuelles qui en découlent (Noë & O'Regan, 2000; Noë et al., 2000)

Armé d'une carte routière, voilà notre homme qui s'escrime à traverser le centre-ville à une heure de pointe. Des véhicules arrivent de toutes parts, le ballet des feux tricolores bat son plein, çà et là des piétons traversent la route, et pour couronner le tout, l'arrosage automatique du parc qui longe la voie parsème son pare-brise de fines gouttes d'eau... Nul ne sera pourtant surpris d'apprendre que notre homme est parvenu à gérer sans encombre la multitude des informations qui ont attiré ses yeux et son attention, et qu'il est arrivé sans difficulté à destination.

Ce travail de thèse s'est intéressé à la question de la perception des scènes visuelles, et s'est plus particulièrement axé sur la représentation des informations visuelles en mémoire lorsque notre exploration visuelle de la scène est dévolue à la réalisation d'une tâche. Reprenons l'exemple de notre automobiliste. Quelles informations a-t-il traitées et suffisamment représentées pour guider sa conduite et arriver à bon port? Toutes les informations présentes dans la scène, de la marque du véhicule qui le précède à la couleur du linge qui sèche sur un balcon ? Les passants, les véhicules et les boutiques qui lui permettent de savoir qu'il circule dans le centre d'une ville ? Les panneaux indicateurs, les marquages au sol, qui lui permettent de se diriger là où il doit aller ? Sachant que les limites attentionnelles et mnésiques ne nous autorisent à prendre en compte qu'une poignée d'éléments seulement, il est donc crucial de s'attacher à comprendre les déterminants du traitement et de la représentation des scènes visuelles dans lesquelles nous évoluons.

Ce travail de thèse propose d'étudier la question des représentations visuelles lorsque notre exploration de la scène est dévolue à la réalisation d'une tâche. Il s'attache plus particulièrement à étudier comment les informations visuelles pertinentes par rapport à la tâche déterminent le contenu de nos représentations visuelles et prennent l'ascendant sur les autres informations visuelles (par exemple les informations sémantiques liées à la

compréhension de la scène), pour être représentées en mémoire et influencer nos comportements.

Dans un premier temps, les arguments théoriques qui témoignent de l'importance que revêtent les demandes de la tâche dans le traitement des scènes visuelles seront présentés. La question de la représentation en mémoire d'une scène visuelle complexe, quand l'exploration visuelle de cette scène est dévolue à la réalisation d'une tâche particulière sera étudiée dans trois chapitres, sous les angles perceptif, attentionnel et mnésique.

Le premier chapitre rendra compte du rôle joué par les connaissances précocement activées dans les processus d'identification et de reconnaissance des scènes visuelles naturelles. Il posera ensuite la nécessité de l'intervention de traitements attentionnels pour que les scènes visuelles puissent être appréhendées et perçues plus en détail. Enfin, puisque les informations visuelles qui seront focalisées par l'attention bénéficieront de traitements appuyés et seront, de fait, perçues plus en détail, ce chapitre témoignera des déterminants de l'orientation attentionnelle au sein d'une scène, et mettra en avant l'influence prépondérante des demandes de la tâche dans le guidage de l'attention. Le chapitre deux, quant à lui, présentera une synthèse des modèles de l'orientation de l'attention qui ont intégré une composante fonctionnelle dans le guidage exogène de l'attention visuelle et permettent de rendre compte du déploiement attentionnel dans des scènes visuelles plus ou moins complexes. Enfin, le chapitre trois s'intéressera plus particulièrement à la question de la représentation des informations visuelles en mémoire. Y seront présentées des études comportementales qui s'intéressent aux déterminants de la représentation des informations visuelles en mémoire et suggèrent que la pertinence des informations visuelles par rapport à la tâche qui nous incombe, mais aussi l'intérêt sémantique de l'information pour comprendre la scène, jouent un rôle important dans la représentation des scènes visuelles. Les résultats de l'ensemble de ces études nous conduiront, compte tenu des limitations de notre système de

traitement, à considérer plusieurs niveaux dans la représentation des informations visuelles. Aussi, dans le chapitre quatre, nous poserons la question de l'architecture du traitement et de la représentation des scènes visuelles complexes lorsque la perception des scènes est contrainte par le fait de devoir réaliser une tâche. Notre hypothèse sera qu'il existe plusieurs niveaux de représentations des informations visuelles, et que le plus haut niveau de ces représentations, qui est aussi le plus accessible par des moyens conscients, et le plus à même d'influencer nos comportements, codera spécifiquement les informations de la scène qui sont pertinentes par rapport à la tâche. Ainsi, dans ce travail de thèse, nous proposerons que le plus haut niveau de représentation des scènes visuelles est un niveau fonctionnel, qui code, de façon dynamique, les éléments de la scène au fur et à mesure qu'ils deviennent immédiatement et directement pertinents et nécessaires à la tâche en cours de résolution.

Les trois chapitres suivants présenteront des études comportementales qui ont été réalisées pour appuyer cette thèse. Le chapitre cinq étudiera comment sont différenciellement représentées les informations visuelles selon leur intérêt sémantique et leur pertinence par rapport à une tâche de recherche visuelle, et suggèrera l'ascendant d'un niveau fonctionnel sur un niveau sémantique des représentations visuelles. Le chapitre six se focalisera plus en détail sur la dynamique représentationnelle des informations pertinentes par rapport à la tâche, au sein de ce niveau fonctionnel de la représentation des scènes. Enfin, le chapitre sept proposera une modélisation de l'influence des demandes de la tâche sur la représentation des informations visuelles.

En dernier lieu, ce travail de thèse discutera de l'apport de ces données expérimentales et des perspectives de recherche qui en découlent.

PREMIÈRE PARTIE :

ASPECTS THÉORIQUES

– CHAPITRE UN –

LA PERCEPTION DE SCÈNES VISUELLES COMPLEXES

Une très grande partie, pour ne pas dire la totalité, de nos comportements s'exprime au sein de scènes visuelles naturelles : le concept de scène visuelle a donc fait l'objet de très nombreuses études en psychologie, en particulier dans le domaine de la perception visuelle.

Pourtant, les scènes visuelles restent définies, faute de mieux, comme des vues sémantiquement cohérentes du monde réel, qui peuvent être nommées et qui, de surcroît, proposent un arrière-plan duquel se détachent de nombreux objets organisés régulièrement dans l'espace (Henderson & Hollingworth, 1999). La difficulté manifeste à en formuler une définition stricte et rigoureuse contraste avec l'extraordinaire capacité de l'être humain à reconnaître et à identifier des scènes visuelles qui s'impriment sur sa rétine, en dépit de larges interférences imputables aux caractéristiques mêmes du système visuel (l'œil est en perpétuel mouvement, l'acuité visuelle est limitée, la vision des couleurs non homogène... Pour plus de détails, voir Bruce & Green, 1993) et aux scènes en elles-mêmes (une scène est porteuse d'une très grande quantité d'informations visuelles distinctes telles les contours, les textures, les couleurs, les mouvements... qui bombardent la rétine en permanence). Ce premier chapitre a pour objectif de mieux comprendre les mécanismes responsables de la perception des scènes visuelles².

² Nous réserverons, tout au long de cette thèse, le terme « scène » aux scènes visuelles qui sont dotées d'une signification, par contraste avec le terme « dispositif visuel » que nous utiliserons en référence à toute scène visuelle simple, dépourvue de signification

1. IDENTIFICATION DE SCÈNES VISUELLES

De nombreuses études ont montré que les scènes visuelles sont reconnues aussi rapidement que des objets présentés isolément (Biederman, Mezzanotte, & Rabinowitz, 1982; Friedman, 1979; Intraub, 1997), et que de très courtes durées de présentation sont suffisantes pour permettre à des sujets humains d'identifier et de labelliser une scène visuelle, c'est-à-dire d'en extraire la signification générale ou le résumé (Intraub, 1981; Irwin, 1992; McCauley, Parmelee, Sperber, & Carr, 1980 ; Potter, Staub, Rado, & O'Connor, 2002; Thorpe, Fize, & Marlot, 1996).

1.1. Résumé et arrangement spatial de la scène

D'une manière générale, le terme « résumé d'une scène » (ou « gist ») fait référence à la signification de la scène et à l'appréhension de son niveau de catégorisation primaire (par exemple, s'il s'agit d'une ville, d'une chambre d'enfant... Tversky & Hemenway, 1983). Il inclut les informations résultant des différents niveaux de traitement de la scène, depuis le traitement précoce des dimensions de bas niveau (couleur, contraste...) jusqu'aux informations dites de haut niveau de la scène (le contexte de la scène et les objets qu'elle englobe, Potter, 1999; Potter, Staub, & O'Connor, 2004; Potter et al., 2002), en passant par les propriétés intermédiaires (surface, volume... Pour une revue, voir Oliva, 2005). Cette connaissance est, de plus, supposée être porteuse d'informations plus générales sur le tracé général de la scène (Hochberg, 1968 ; Intraub, 1997; Simons, 1996), l'arrangement spatial (Sanocki, 2003; Schyns & Oliva, 1994; Simons, 1996) et/ou la couleur (Oliva & Schyns, 2000) des objets qui la composent. Le résumé de la scène peut être appréhendé extrêmement rapidement, en une seule fixation (Biederman, 1972; Intraub, 1981), voire en des durées bien

plus courtes (20ms) que celle d'une fixation oculaire³ (Bacon-Macé, Macé, Fabre-Thorpe, & Thorpe, 2005; Delorme, Richard, & Fabre-Thorpe, 2000; Fabre-Thorpe, Delorme, Marlot, & Thorpe, 2001; Thorpe et al., 1996; VanRullen & Thorpe, 2001a), mais il est alors nécessaire que le traitement de l'image puisse perdurer au moins 150 à 200 ms après que l'image a disparu du champ visuel pour que cette information puisse se consolider (Thorpe et al., 1996; VanRullen & Thorpe, 2001a). En effet, lorsque plusieurs scènes visuelles sont présentées en série à un rythme supposé simuler les interruptions saccadiques (173 ms de présentation par image), les participants échouent en très grande majorité à un test de reconnaissance des scènes, ce qui suggère qu'une scène visuelle peut être oubliée aussi rapidement qu'elle a été appréhendée lorsque son traitement est court-circuité par le traitement de l'image suivante (Intraub, 1981; Potter, 1976 ; Potter et al., 2004; Potter et al., 2002). Aussi, la réussite des processus d'identification des scènes visuelles présentées très brièvement requerrait l'intervention de traitements visuels et attentionnels supplémentaires, ainsi que l'accumulation continue d'informations visuelles, consécutivement à l'activation du résumé de la scène (Bacon-Macé et al., 2005 ; Jiang, 2004). Par ailleurs, la rapidité des processus de reconnaissance d'une scène visuelle est telle que, dans une tâche de catégorisation de type go/no go⁴, les participants parviennent à déterminer avec des taux de réussite importants (95%) si des images présentées extrêmement brièvement (20ms) contiennent une cible particulière (par exemple un animal), que les scènes soient présentées en couleur (Thorpe et al., 1996) ou en noir et blanc (Delorme et al., 2000), que la cible appartienne à une catégorie biologique ou non (VanRullen & Thorpe, 2001a), et cela sans que la performance soit améliorée par un entraînement (Fabre-Thorpe et al., 2001), ou mobilise des ressources

³ La durée moyenne d'une fixation oculaire sur une scène visuelle naturelle est estimée aux alentours de 300 ms (Henderson, 2003; O'Regan, Deubel, Clark, & Rensink, 2000)

⁴ Les participants doivent relâcher un bouton si la scène présentée contient la cible préalablement définie, par exemple un animal (go), ou, au contraire, maintenir le bouton pressé si la scène ne contient pas la cible (no go)

attentionnelles (Li, VanRullen, Koch, & Perona, 2002). Les conditions de la réussite de cette tâche indiquent que le résumé de la scène est très probablement appréhendé très précocement durant le traitement visuel de la scène (Bar, 2004; Biederman, 1972 ; Irwin & Yeomans, 1986; Oliva, 2005 ; Potter, 1976, 1993, 1999; Potter et al., 2004; Potter et al., 2002 ; Wolfe, 1998), probablement à partir du résultat des traitements préattentifs, qui calculent massivement, et en parallèle, chacune des dimensions de bas niveau de la scène (couleur, orientation, contraste, mouvement... Treisman, 1998; Treisman & Gelade, 1980).

1.2. Le schéma de scène

Un second type de connaissances, plus conceptuel que le résumé de la scène, et connu sous le terme « schéma de scène »⁵, serait également activé dès les premiers regards sur les scènes visuelles (Biederman, 1972; Biederman et al., 1982; Potter, 1993). Les schémas de scène sont des représentations prototypiques de chaque catégorie de scènes visuelles, stockées en mémoire à long terme, qui s'apparentent à un inventaire des objets susceptibles d'être présents dans telle ou telle catégorie de scènes visuelles (*eg.*, on trouve, typiquement, un lit et des jouets dans une chambre d'enfant). Ils contiennent, de plus, des informations sur les régularités spatiales associées à ces objets et les localisations relatives des objets les uns par rapport aux autres (*eg.*, les oreillers sont posés à la tête du lit), ainsi que des connaissances plus génériques relatives aux contraintes physiques et à la plausibilité biologique (*eg.*, un lit ne flotte pas dans les airs, Biederman et al., 1982; Mandler & Ritchey, 1977). Lorsqu'une scène visuelle s'imprime sur notre rétine, l'activation, quasi immédiate, du schéma de la scène en mémoire détermine le traitement des objets affichés par la scène : les objets

⁵ Ou « schéma perceptif »

sémantiquement cohérents avec le schéma de scène sont traités rapidement et normalisés en valeurs par défaut, alors que les objets sémantiquement incohérents bénéficient d'un traitement plus appuyé (Friedman, 1979) suivi d'une consolidation mnésique, qui conduit à l'enrichissement du stock de connaissances sur les scènes en mémoire. Les schémas de scène jouent donc, de fait, un rôle fondamental dans la perception et la mémorisation des scènes visuelles naturelles. Par exemple, des recherches ont montré que des participants soumis à un test de mémoire tendent à rappeler ou à reconnaître, de façon quasi systématique, davantage d'éléments que la scène originale n'en contenait réellement. En effet, lorsque l'on présente aux participants plusieurs prises de vues d'une même scène, prises de vue qui ne diffèrent entre elles que par la focale de l'objectif (*ie.*, qui proposent une vue soit plus restreinte, soit plus élargie que celle de la scène originale), les participants tendent à désigner la vue la plus élargie comme correspondant à la photographie antérieurement perçue (Intraub, Bender, & Mangels, 1992; Intraub & Bodamer, 1993; Intraub, Gottesman, & Bills, 1998). Ce phénomène, connu sous le terme d'« extension de frontières⁶ », est extrêmement robuste⁷ et témoigne de l'influence des schémas perceptifs, stockés en mémoire à long terme, sur la perception de scène. Tout se passe comme si la scène originalement présentée entraînait très rapidement l'activation d'un schéma perceptif en mémoire, schéma qui contient des connaissances sur les éléments supposés être présents dans la scène. Lorsque les participants sont ensuite confrontés à un test de mémoire, ils utilisent le schéma de scène préalablement activé pour rappeler les informations contenues dans la scène. Ce faisant, ils rappellent non

⁶ En anglais : « boundaries extension »

⁷ Il se retrouve, que les durées de présentation de la scène soient longues (15 s) ou courtes (250 ms) (Intraub et al., 1998), que la procédure expérimentale utilise des tests de reconnaissance (Intraub et al., 1992) ou de reproduction de la scène par le dessin (Intraub & Bodamer, 1993), et résiste à la mise en garde des participants contre cette tendance à reconnaître comme cible une photographie présentant des frontières plus larges que celles de la cible (Intraub & Bodamer, 1993)

seulement les informations présentes dans la scène originale mais également les informations absentes de la scène mais présentes dans le schéma de cette scène.

Bien que les connaissances relatives au résumé, à l'arrangement et au schéma d'une scène visuelle soient suffisantes pour comprendre, labelliser, identifier et encoder très rapidement une scène visuelle, elles n'en permettent pas le traitement détaillé. En fait, la perception du détail des scènes visuelles va nécessiter que soient, par la suite, engagés des processus attentionnels qui permettront non seulement à l'identification rapide des scènes de résister aux interférences mais surtout de rendre possible le traitement et la représentation en mémoire des objets visuels qui composent les scènes visuelles (Henderson, Williams, Castelhana, & Falk, 2003).

En fait, le rôle joué par l'attention visuelle dans la perception de scènes s'exprime de multiples manières, et ses effets s'observent autant du point de vue comportemental que neuronal (Spitzer, Desimone, & Moran, 1988). D'une manière générale, l'attention permet de caractériser et de distinguer les objets visuels de la scène (Scholl, 2001; Treisman, 1998 ; Wolfe, 1994) et favorise leur représentation consciente (Dennet, 1993; Rensink, 2001), ce qui fait d'elle la candidate idéale de l'interface entre la vision et la cognition (Pylyshyn, 1999). De manière alternative, les contrecoups engendrés par la mobilisation de l'attention sur tel ou tel aspect de la scène sont considérables, et soulignent la faible efficacité de nos capacités perceptives lorsqu'elles sont amputées des traitements attentionnels (Chun & Marois, 2002). Par exemple, lorsque le focus de l'attention est engagé ailleurs, les informations visuelles portées par une scène voient leurs chances d'être traitées et/ou représentées en mémoire et/ou accessibles de manière consciente chuter drastiquement, phénomènes connus sous les termes de cécité inattentionnelle (Mack & Rock, 1998; Most, Clifford, & Simons, 2005; Simons, 2000a; Simons & Chabris, 1999), de cécité au changement (O'Regan, 2001; Simons &

Ambinder, 2005; Simons & Levin, 1997; Simons & Rensink, 2005) et de clignement attentionnel (Chun & Potter, 1995; Joseph, Chun, & Nakayama, 1997; Luck, Vogel, & Shapiro, 1996). Aussi, l'étude de la perception de scènes visuelles complexes ne peut s'affranchir de l'étude du déploiement de l'attention au sein d'une scène visuelle, et les facteurs de l'orientation de l'attention aideront à identifier les éléments de la scène qui seront les plus à même de bénéficier de traitements attentionnels soutenus et donc d'être consciemment perçus.

2. ATTENTION ET PERCEPTION DES SCÈNES VISUELLES

Compte tenu du rôle central joué par l'attention et la pluralité de ses effets dans le domaine de la cognition humaine et animale, nous nous attacherons, dans cette section, à distinguer les facteurs responsables de l'orientation attentionnelle au sein d'une scène visuelle, laquelle permettra ensuite l'extraction et le traitement détaillé de certaines informations visuelles affichées par la scène en vue de leur représentation en mémoire. Au lecteur consciencieux, nous conseillerons de consulter les quelques très bonnes revues de littérature sur la question attentionnelle (Egeth & Yantis, 1997; Klein, Kingstone, & Pontefract, 1992; Klein & Shore, 2000; Pashler, 1998; Posner, 1980; Scholl, 2001 ; Wright, 1998)⁸. Plus particulièrement, nous nous attacherons, dans cette section, à distinguer les facteurs responsables de l'orientation involontaire et automatique (ou exogène) de l'attention d'une part et de l'orientation intentionnelle (ou endogène) de l'attention d'autre part, tout en traitant de manière plus ou moins indifférenciée de l'attention explicite, qui peut être

⁸ Il est intéressant de remarquer, au passage, que la multitude des revues de questions consacrées à l'attention témoigne de l'importance que revêtent les traitements attentionnels dans la cognition humaine

appréhendée à partir des orientations manifestes de la tête et du regard, et de l'attention implicite, qui doit être inférée à partir des performances observées (Klein et al., 1992; Klein & Shore, 2000; Posner, 1980).

Plusieurs raisons motivent ce choix. En premier lieu, l'attention implicite, tout comme l'attention explicite, conduit au traitement approfondi et à la représentation en mémoire de l'information visuelle (Posner, 1980), qui constituent véritablement le fil rouge de ce travail de thèse : il ne nous est donc pas nécessaire de les distinguer. Ensuite, il a été montré que même si les déplacements attentionnels précédaient les déplacements oculaires (Posner, 1980), les mouvements oculaires ne constituaient pas moins de bons indicateurs des traitements attentionnels et cognitifs (Gilchrist, Brown, & Findley, 1997; Liversedge & Findley, 2000), ce qui n'est guère surprenant compte tenu du fait que certaines aires cérébrales connues pour être le siège de l'attention visuelle, à savoir le cortex pariétal postérieur et plus particulièrement l'aire intrapariétale (Gottlieb, Kusunoki, & Goldberg, 1998), sont aussi à l'origine de la programmation des saccades oculaires (Snyder, Batista, & Andersen, 1997), et que l'œil, tout comme l'attention, tend à se concentrer sur des objets plutôt que sur des régions lorsque nous sommes confrontés à des scènes visuelles (Henderson, Weeks, & Hollingworth, 1999; Scholl, 2001). Pour toutes ces raisons, nous traiterons donc, dans la section qui va suivre, des facteurs de l'orientation exogène et endogène de l'attention, qu'elle soit implicitement ou explicitement orientée.

2.1. L'attention visuelle s'oriente vers certains stimuli visuels de manière exogène

Il est classiquement admis qu'il existe des facteurs exogènes (ou bottom-up, c'est-à-dire dirigés par le stimulus) qui attirent irrésistiblement l'attention sur tel ou tel item de la scène,

et peuvent en entraîner, par conséquent, le traitement détaillé. On parle donc d'attention exogène en référence aux déplacements attentionnels occasionnés par la saillance d'un item ou d'un évènement de façon involontaire et irréprensible. Ce type de guidage exogène de l'attention a surtout été mis en évidence dans des tâches de recherche visuelle, dans lesquelles il s'agit de déterminer si une cible est présente parmi un ensemble de distracteurs. Ce genre de tâche est considéré par Wolfe (1994) comme une bonne approximation des tâches de recherches visuelles ordinairement engagées dans la vie de tous les jours, quand il s'agit de mettre la main sur un trousseau de clé, de reconnaître un visage connu dans une foule en mouvement, etc... Dans les tâches de recherche visuelle en laboratoire, où il est d'usage d'utiliser des dispositifs visuels extrêmement simples et affranchis de toute signification, on considère que si la performance (mesurée en termes de réponses correctes ou de temps de réponse) ne dépend pas de la taille du dispositif⁹, alors les items ont été traités en parallèle, et de manière préattentive (Wolfe, 1994). Plusieurs dimensions basiques des scènes visuelles ont pu être ainsi désignées comme étant traitées de manière préattentive : la couleur, l'orientation, le mouvement, et la taille du stimulus, parmi d'autres (Francolini & Egeth, 1980; Treisman, 1998; Treisman & Gelade, 1980; Wolfe, 1994; Wolfe & Horowitz, 2004). Par exemple, lorsque l'on demande à des participants de déterminer le plus rapidement possible si une cible rouge est présente parmi un ensemble de distracteurs verts, ils répondent aussi rapidement à la présence de la cible quel que soit le nombre de distracteurs alentours (l'on retrouve, bien entendu, des résultats similaires pour une cible qui se distinguerait des distracteurs par une orientation, un mouvement ou une taille particulière). Ainsi, ces dimensions basiques préattentives serviraient de base au guidage attentionnel au sein des scènes visuelles, de manière purement exogène (Duncan & Humphreys, 1989 ; Treisman & Gelade, 1980; Wolfe,

⁹ On parle également du nombre de distracteurs pour exprimer l'effet de la taille du dispositif

1994; Wolfe & Horowitz, 2004). Cependant, comme le font remarquer Jonides et Yantis (1988), ce n'est pas parce que des items sont facilement localisés sur la base de la saillance d'une de leurs dimensions que cette dimension attire automatiquement l'attention. De plus, dans la mesure où les recherches ayant conduit à des telles conclusions ont inséré des cibles uniques au sein de distracteurs semblables entre eux et très dissemblables à la cible (par exemple une cible rouge parmi des distracteurs verts), il est difficile de pouvoir attribuer la constance de la performance de recherche visuelle quelle que soit la taille du dispositif à l'effet d'une des dimensions basiques de la cible (couleur, orientation...) ou au fait que la cible constitue un singleton¹⁰ (unique élément rouge, unique élément à orientation verticale...). Aussi, certains auteurs ont préféré considérer le degré de ressemblance / dissemblance qui distingue la cible des distracteurs pour rendre compte du guidage exogène de l'attention (Braun & Sagi, 1990; Duncan & Humphreys, 1989). Ainsi, dans des tâches de recherche visuelle utilisant des paradigmes aussi variés que le singleton additionnel¹¹ (Theeuwes, 1994), la capture oculaire¹² (Brockmole & Henderson, sous presse-a; Theeuwes, Kramer, Hahn, & Irwin, 1998), la recherche de la dimension non pertinente¹³ (Jonides & Yantis, 1988) parmi d'autres (pour une revue, voir Simons, 2000a), il a été montré que des items qui se différencient suffisamment des autres pour être considérés comme des singletons du point de vue de la couleur (Theeuwes, 1994; Theeuwes, Kramer, & Kingstone, 2004), de

¹⁰ Nous appellerons « singleton » tout élément qui est localement unique du point de vue d'une dimension perceptive (couleur, orientation...)

¹¹ Dans le paradigme du singleton additionnel les participants doivent réaliser une tâche de recherche visuelle, et un item unique (un singleton) apparaît dans la scène : on dit qu'il y a capture exogène de l'attention si la performance est ralentie par l'apparition du singleton additionnel

¹² Dans le paradigme de la capture oculaire, les participants doivent réaliser une tâche de recherche visuelle, et un item unique (un singleton) apparaît dans la scène : on dit qu'il y a capture exogène de l'attention si on observe une saccade oculaire en direction du singleton additionnel

¹³ Dans le paradigme de recherche de la dimension non pertinente, un singleton apparaît dans la scène soit en position de cible, soit en position de distracteur : on considère qu'il y a capture exogène de l'attention si la réponse des participants est plus rapide lorsque la cible est le singleton que lorsqu'elle ne l'est pas.

la forme (Theeuwes, 2005), du mouvement (McLeod, Driver, & Crisp, 1988), ou de l'orientation (Treisman, 1998), attirent l'attention de manière irrésistible, qu'ils soient ou non la cible de la tâche de recherche visuelle. Les singletons constituent donc des déterminants du guidage exogène de l'attention dans un dispositif visuel.

Enfin, l'apparition soudaine d'un item dans le dispositif visuel serait aussi un facteur important de l'orientation exogène de l'attention visuelle (Brockmole & Henderson, sous presse-a; Jonides & Yantis, 1988; Peterson & Kramer, 2001; Theeuwes, 1994; Theeuwes et al., 1998), ce qui n'est guère surprenant compte tenu de fait que notre aptitude à nous orienter rapidement vers un stimulus introduit soudainement dans notre champ visuel a très certainement favorisé la survie de l'espèce humaine¹⁴. Il semblerait donc que nous devions considérer que le déploiement de l'attention au sein d'une scène, indispensable au traitement approfondi des objets de la scène, est contraint par la saillance perceptive des éléments de la scène (singletons et apparitions brusques) : les éléments les plus perceptivement saillants de la scène capturent l'attention de manière automatique et irrésistible et ont donc de bonnes chances d'être représentés en mémoire, indépendamment des buts et attentes des observateurs.

Or, une majeure partie des arguments en faveur d'une orientation exogène de l'attention dans une scène visuelle provient de recherches dans lesquelles l'élément saillant (le singleton ou l'élément introduit soudainement dans le dispositif visuel) partage des caractéristiques communes avec la cible. Par exemple, la cible est caractérisée comme étant un singleton de forme (seul carré parmi des cercles) et l'élément saillant est un singleton de couleur (seul item rouge parmi des items verts). Il est donc possible que les effets observés sur les performances des participants, interprétés comme le reflet de la capacité de certains items à attirer

¹⁴ L'attraction de l'attention par un élément qui apparaît brièvement dans la scène peut, en fait, résulter de la soudaineté de l'apparition ou du fait que l'objet apparu dans la scène est nouveau. Il a été montré que si l'on supprime la soudaineté de l'insertion d'un nouvel objet, celui-ci n'est plus en mesure d'attirer l'attention exogène (Brockmole & Henderson, sous presse-a). C'est donc bien l'apparition brusque, et non la nouveauté, d'un stimulus qui est un déterminant de l'orientation exogène de l'attention.

l'attention de façon exogène, résultent tout simplement de l'utilisation stratégique de connaissances sur la cible, de type « puisque la cible est un singleton, je cherche un singleton ». Dans ce cas, l'orientation supposée exogène de l'attention sur les items saillants serait, en définitive, le fait d'une orientation *endogène* de l'attention (Yantis & Egeth, 1999).

Afin de distinguer les effets exogènes des effets endogènes sur le déploiement attentionnel, Ludwig et Gilchrist (2002) ont examiné la propension de singletons de couleur introduits dans un dispositif visuel à attirer l'attention, selon qu'ils partagent les caractéristiques de la cible (dans ce cas l'item est semblable à la cible) ou non (dans ce cas l'item est dissemblable à la cible). La cible est un item rouge parmi des distracteurs verts. Le principe de l'expérience, qui utilise un paradigme de recherche visuelle avec singleton additionnel, est basé sur l'influence des demandes de la tâche sur l'orientation exogène de l'attention vers l'item additionnel, mesurée via les mouvements oculaires et les temps de réponse. Si les singletons de couleurs sont réellement des facteurs de l'orientation exogène de l'attention, alors la réponse à la cible devrait être altérée par l'apparition d'un singleton, que ce singleton soit semblable ou dissemblable à la cible. Par contre, si les singletons de couleurs ne sont pas systématiquement des facteurs de l'orientation exogène de l'attention et que leur influence sur le déploiement attentionnel résulte plutôt de l'application d'une stratégie contingente aux demandes de la tâche (trouver un singleton de couleur), alors la réponse à la cible devrait être altérée uniquement par l'apparition d'un singleton qui lui est semblable. Les résultats montrent que les temps de réponse à la cible sont ralentis uniquement lorsque le singleton introduit dans le dispositif visuel est semblable à la cible. De manière comparable, les patterns de mouvements oculaires révèlent que les yeux sont davantage attirés par le singleton additionnel lorsqu'il est semblable à la cible que lorsqu'il est dissemblable à la cible. Ce premier résultat suggère fortement que ce type de guidage attentionnel, supposé être exogène, peut en fait être assujéti à la tâche et émerger de l'utilisation stratégique des

connaissances sur les caractéristiques de la cible. Dans ce cas, l'attention ne serait réellement attirée par des items saillants que lorsqu'ils partageraient les caractéristiques de la cible, ce qui conforterait l'hypothèse selon laquelle il existe peu de facteurs qui soient purement exogènes et affranchis du joug de la commande « top down ».

Bien entendu, d'autres recherches soulignent l'influence inhibitrice des facteurs endogènes (demandes de la tâche, connaissances sur la cible...) sur les facteurs exogènes de l'attention (Folk, Remington, & Johnston, 1992; Peterson & Kramer, 2001 ; Yantis & Egeth, 1999). Par exemple, lorsqu'une consigne informe des participants sur le fait qu'un item saillant par une de ses dimensions (par exemple, un singleton de couleur ou d'orientation) ne correspondra que très rarement à la cible de la tâche de recherche visuelle (cette consigne a pour conséquence de réduire la saillance endogène des singletons sans en réduire la saillance exogène), les temps de réponse à la présence de la cible ne varient guère, que la cible corresponde à l'item saillant ou non (Yantis & Egeth, 1999). Or, si l'attention avait été irrésistiblement attirée par l'élément saillant, la réponse à la cible aurait dû être plus rapide lorsque la cible correspond à l'élément saillant que lorsqu'elle ne lui correspond pas. Ce résultat indique, là encore, que l'orientation exogène de l'attention s'avère dépendante de processus top-down. Bien entendu, la modulation de l'orientation exogène de l'attention par des facteurs top-down n'est pas limitée aux tâches de recherche visuelle. Par exemple, dans une tâche de poursuite visuelle, où il s'agit de suivre du regard un petit nombre de ronds noirs se déplaçant aléatoirement dans un dispositif visuel, les participants sont dans l'incapacité de percevoir qu'un élément interférent (une croix) traverse le dispositif de part en part, à moins que cet élément ne partage *au moins une* des caractéristiques de la cible (par exemple, la couleur noire), et cela en dépit du fait que cet élément interférent soit un singleton de couleur, de mouvement, et apparaisse soudainement dans le champ visuel des participants (Most et al., 2005). Ce résultat indique, là encore, que la tâche a engendré une préparation attentionnelle

aux propriétés de la cible, et que la présence d'un distracteur, même extrêmement saillant du point de vue perceptif, est ignorée des traitements attentionnels dès lors qu'il se différencie de la cible. Plus généralement, tous ces résultats sont en accord avec la théorie avancée par Folk et collaborateurs (Folk et al., 1992), pour qui l'orientation exogène de l'attention est contingente des demandes de la tâche. Ainsi, la capture attentionnelle ne serait jamais purement exogène, puisque le déploiement involontaire de l'attention sur un item saillant mais non pertinent par rapport à la tâche ne s'observerait que lorsque cet item présente des propriétés de la cible auxquelles les participants s'étaient préparés à répondre. Ainsi, l'allocation supposée exogène de l'attention sur les items saillants d'un dispositif visuel résulterait, en fait, de l'utilisation stratégique des connaissances sur les caractéristiques de la cible, et dépendrait donc des demandes de la tâche.

Enfin, il est intéressant de relever que l'annihilation de l'influence des facteurs exogènes par des facteurs endogènes du déploiement attentionnel peut être également dépendante de la charge mentale mobilisée, et donc de la quantité de ressources cognitives disponibles. Ainsi, des items perceptivement saillants (singletons de couleur, d'orientation, ou apparitions soudaines) attirent d'autant moins l'attention de manière exogène qu'une bonne partie, voire la totalité, des ressources attentionnelles est par ailleurs mobilisée par une tâche concomitante (Boot, Brockmole, & Simons, sous presse; Joseph et al., 1997; Yantis & Johnston, 1990; Yantis & Jonides, 1990).

Considérée dans son ensemble, cette série de résultats semble suggérer que, bien que l'attention puisse être guidée de manière purement exogène par certains stimuli particulièrement saillants et pertinents du point de vue de la survie de l'espèce, *ie.*, les apparitions soudaines dans le champ visuel (Jonides & Yantis, 1988; Yantis & Egeth, 1999), ou le mouvement (Brockmole & Henderson, sous presse-a; Yantis & Egeth, 1999), le

déploiement de l'attention au sein d'une scène visuelle serait tout de même dépendant des processus endogènes, eux-mêmes dépendants de la tâche à réaliser.

2.2. La perception des scènes dépend de l'orientation endogène de l'attention visuelle

Le guidage volontaire de l'attention vers certaines informations de la scène permet de leur allouer les ressources nécessaires pour qu'elles soient traitées de manière plus détaillée et suffisamment représentées en mémoire pour que nous y ayons accès consciemment. Ainsi, lorsque nous voulons percevoir plus en détail certains objets de la scène, nos connaissances sur la scène (le résumé, le tracé et le schéma de scène) vont guider le déploiement de l'attention vers ces objets (Henderson, 2003 ; Oliva, 2005; Rensink, 2000a; Torralba, 2003), en facilitant notamment les processus d'apprentissage implicite qui sous-tendent l'indiciage contextuel¹⁵.

L'indiciage contextuel désigne l'apprentissage et l'utilisation implicites de la covariation régulière d'un contexte particulier avec la position d'une cible, qui permettent de guider efficacement l'attention vers la cible (Brockmole & Henderson, sous presse-b; Chun & Jiang, 1998; Chun & Nakayama, 2000; Chun & Wolfe, 2001; Peterson & Kramer, 2001). Ce phénomène entre particulièrement en jeu dans la perception de scènes visuelles naturelles, puisque les scènes naturelles ont des structures relativement stables et régulières¹⁶ (Chun, sous presse ; Schyns & Oliva, 1994; Torralba & Oliva, 2003) et que, par ailleurs, nous

¹⁵ « Contextual cueing »

¹⁶ Par exemple, une scène urbaine contient régulièrement des routes sur lesquelles circulent des véhicules, lesquelles sont encadrées par des immeubles qui se dressent vers le ciel et des trottoirs arpentés par des passants, les panneaux de signalisations y sont placés à des endroits stratégiques...

disposons d'informations sur la localisation relative des objets prototypiques par le biais des schémas de scènes (Friedman, 1979). Il existe donc des régularités dont le système cognitif peut tirer parti pour déterminer efficacement les zones de la scène susceptibles de contenir des informations intéressantes et qu'il serait donc utile de traiter, et, au contraire, les zones de la scène qui ont peu de chance d'être informatives et vers lesquelles il serait vain d'orienter les traitements oculaires et attentionnels (Chun, 2000, sous presse). Ainsi, par le biais de l'indiciage contextuel, les connaissances activées très précocement par la présentation d'une scène visuelle, à savoir le résumé, le tracé général, et le schéma de la scène vont déterminer quels objets de la scène seront ignorés et lesquels bénéficieront de traitements ultérieurs. Quelles seront ces informations visuelles préférentiellement traitées par l'attention ?

Lorsque l'on laisse libre l'exploration des scènes visuelles (autrement dit, lorsque les participants sont affranchis de toute consigne autre que celle de regarder la scène), il a été montré une propension de l'œil à se poser de façon systématique et récurrente sur certains objets de la scène, et à en ignorer d'autres. Fait surprenant, cette caractéristique de l'exploration visuelle de la scène ne varie pas en fonction de la durée de présentation de la scène, et le pattern des mouvements oculaires observé chez différents sujets reste très similaire, que la scène soit présentée brièvement ou plus longuement, comme le montre la Figure 1 (O'Regan et al., 2000; Yarbus, 1967).

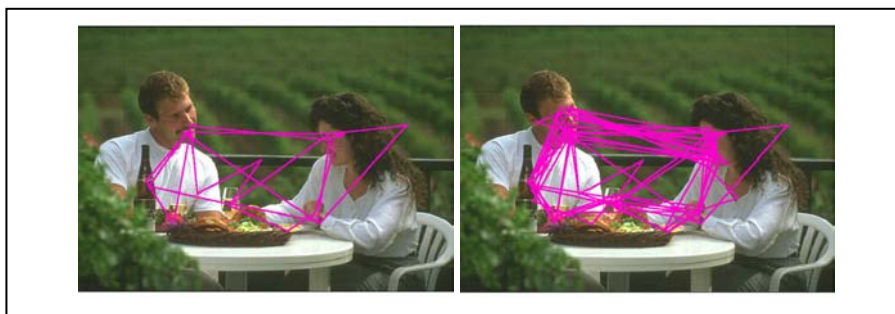


Figure 1 Illustration des mouvements oculaires enregistrés pendant l'exploration d'une scène visuelle naturelle (O'Regan et al., 2000). Les points indiquent les fixations oculaires et les segments les saccades oculaires : le pattern des mouvements oculaires reste constant, que la scène soit présentée pendant 3 secondes (à gauche) ou 40 secondes (à droite)

Ainsi, l'exploration visuelle d'une scène est répétitive et stéréotypée, l'œil et/ou l'attention tendant à se concentrer sur les objets de la scène qui sont « nécessaires et essentiels à la perception » (Yarbus, 1967, p175). Il existe en effet une relation entre la densité des fixations sur les objets de la scène et la quantité d'informations dont ils sont porteurs : les objets de la scène les plus sémantiquement informatifs (c'est-à-dire les objets les plus importants pour la compréhension et l'interprétation de la scène) sont aussi la cible du plus grand nombre de fixations oculaires (Mackworth & Morandi, 1967; O'Regan et al., 2000). Ce résultat suggère que ces objets, dits d'intérêt central, reçoivent un traitement attentionnel privilégié en vertu de leur saillance sémantique. Cependant, il est difficile de faire précisément la part des effets relatifs à la saillance perceptive et ceux relatifs à la saillance sémantique des objets désignés comme étant d'intérêt central, puisque ces derniers peuvent également être saillants du point de vue perceptif (ils sont souvent situés au premier plan et sont donc plus gros et plus lumineux que les autres objets, dits d'intérêt marginal). Par exemple, dans l'expérience de Mackworth et Morandi (1967), les régions cotées les plus informatives peuvent avoir reflété les régions les plus perceptivement saillantes (*ie.*, les régions de la scène les plus visuellement chargées, en matière de contrastes, de contours, de textures). Aussi, afin d'étudier le rôle des facteurs sémantiques sur l'exploration des scènes visuelles tout en contrôlant la saillance perceptive des objets, plusieurs auteurs ont étudié l'exploration des scènes visuelles complexes en fonction de la cohérence des objets de la scène avec leur contexte. L'idée sous-jacente à ces recherches est simple : si le contexte et le schéma perceptif de la scène entraînent tous deux le traitement des objets de la scène qui répondent à un critère sémantique, les objets les plus informatifs du point de vue de la signification de la scène devraient bénéficier de traitements visuels et attentionnels approfondis tandis que les objets peu informatifs devraient être ignorés des traitements visuels et attentionnels. Ainsi, un même objet (*eg.*, un microscope) devrait être différenciellement

traité selon qu'il est intégré dans un contexte cohérent (*eg.*, un laboratoire) ou incohérent (*eg.*, une cuisine), alors même que la saillance perceptive de l'objet est inchangée et donc ne peut rendre compte des différences de traitement. De nombreuses études ont ainsi pu témoigner de l'influence robuste du contexte sur l'identification des objets visuels, qui s'exprime dans différents paradigmes expérimentaux, tels que l'enregistrement des mouvements oculaires (De Graef, Christiaens, & d'Ydewalle, 1990; De Graef, Lauwereyns, & Verfaillie, 2000; Loftus & Mackworth, 1978), les tests de mémoire (Biederman, 1972; Friedman, 1979), les paradigmes de détection des changements (Hollingworth & Henderson, 2000), les tests de détection (Biederman et al., 1982; Li et al., 2002) ou de dénomination d'objets (Boyce & Pollatsek, 1992). Ces différentes études indiquent de façon unanime que les objets sont d'autant plus rapidement reconnus et identifiés qu'ils sont enchâssés dans un contexte cohérent : une chaise est plus rapidement reconnue si elle est insérée dans une salle à manger que si elle est insérée dans une forêt (pour une revue, voir De Graef, 1992; Henderson & Hollingworth, 1999). De plus, les objets sémantiquement cohérents avec le contexte de la scène attirent l'attention plus précocement que les objets sémantiquement incohérents (De Graef, 1998; De Graef et al., 2000), ce qui tendrait à indiquer que l'activation précoce de connaissances conceptuelles et contextuelles instancierait le traitement des objets cohérents avant le traitement des objets incohérents. Ce résultat peut, de prime abord, sembler en désaccord avec d'autres résultats. En effet, aucune des recherches mettant en jeu l'étude des mouvements oculaires n'a pu conclure à une primauté des saccades oculaires sur des objets sémantiquement consistants avec le contexte, ces recherches étant parvenues, tout au plus, à un niveau de significativité tendanciel de leurs résultats : les trois premières fixations au sein d'une scène visuelle sembleraient se concentrer davantage sur les objets cohérents que sur les objets incohérents avec le contexte (De Graef et al., 1990; Henderson & Hollingworth, 1999). La mollesse de ces résultats inciterait donc à traiter avec précaution de l'attraction de

l'attention par les objets sémantiquement cohérents avec le contexte. Toutefois, le traitement précoce des objets contextuellement cohérents peut être légitimé par les résultats obtenus par De Graef et collaborateurs (De Graef et al., 1990) dans une étude où le traitement d'un objet-amorce est interrompu par l'apparition soudaine, dans le champ visuel des participants, d'un objet en vibration. Ces résultats indiquent que les premières fixations oculaires sur les objets en vibration sont davantage différées dans le temps lorsqu'ils sont sémantiquement cohérents que lorsqu'ils sont sémantiquement incohérents avec le contexte, et que, de surcroît, le traitement de l'objet-amorce subit moins d'interférences (en termes de durée de la première fixation et nombre de refixations sur l'objet) quand l'objet en vibration est cohérent que lorsqu'il est incohérent avec le contexte de la scène. Ces résultats sont interprétés par De Graef et ses collaborateurs comme signant la capacité du système cognitif à pouvoir extraire de manière périphovéale davantage d'informations sur les objets en vibration lorsqu'ils sont cohérents que lorsqu'ils sont incohérents avec le contexte. Ces conclusions permettent alors d'expliquer pourquoi la capacité des objets contextuellement cohérents à attirer précocement l'attention ne se retrouve que partiellement dans les expériences utilisant les mouvements oculaires comme mesure de l'orientation attentionnelle. En effet, les objets cohérents attirent l'attention précocement sans n'attirer les mouvements oculaires tout aussi précocement parce qu'ils peuvent être traités de manière périphovéale. Ainsi, comme ils ne nécessitent pas forcément un traitement fovéal, leur influence ne se manifeste que peu sur les mouvements oculaires. Enfin, notons que si les premiers traitements attentionnels sur la scène se concentrent sur les objets cohérents, les traitements ultérieurs se concentrent, quant à eux, sur les objets incohérents. En effet, les objets incohérents sont la cible de fixations oculaires ultérieures plus longues et plus nombreuses que les objets cohérents, ce qui est expliqué par le fait que les objets incohérents sont plus informatifs que les objets cohérents (pour une revue, voir Henderson & Hollingworth, 1999).

Lorsqu'une scène visuelle ne contient que des objets congrus, il est nécessaire pour le système cognitif de donner une priorité attentionnelle à certains objets de la scène. Les meilleurs candidats à l'orientation sélective de l'attention sont les objets d'intérêt central, qui sont les objets de la scène porteurs de la plus grande quantité d'informations sur la signification de la scène (O'Regan et al., 2000; Yarbus, 1967). Ces objets d'intérêt central sont les plus à même d'attirer l'attention très précocement (Mathis, 2002), et d'être régulièrement la cible de nouvelles fixations, par le biais du « priming of popout », qui facilite le redéploiement de l'attention et des yeux sur les régions de la scène les plus pertinentes pour le système cognitif (Maljkovic & Nakayama, 2000). Ainsi, les objets de la scène d'intérêt central pour l'appréhension d'une scène visuelle semblent être la cible privilégiée des traitements attentionnels, en condition de vision passive, c'est-à-dire quand il n'y a pas de tâche à réaliser sur la scène (O'Regan et al., 2000). Le guidage de l'attention au sein d'une scène est donc un processus actif, visant à extraire, de façon hautement sélective, les informations de la scène qui seront d'intérêt pour l'observateur. Toutefois, nous allons voir que ces informations d'intérêt vont quelque peu différer en condition de vision active, c'est-à-dire quand l'exploration visuelle est dévolue à la satisfaction d'une tâche. En effet, en sus d'être guidée par la sémantique de la scène, l'attention peut être guidée de manière endogène par les buts des observateurs. Une des premières recherches à avoir mis en évidence l'influence des demandes de la tâche sur l'exploration visuelle d'une scène est très certainement celle de Yarbus (1967), dont les résultats sont illustrés dans la Figure 2.

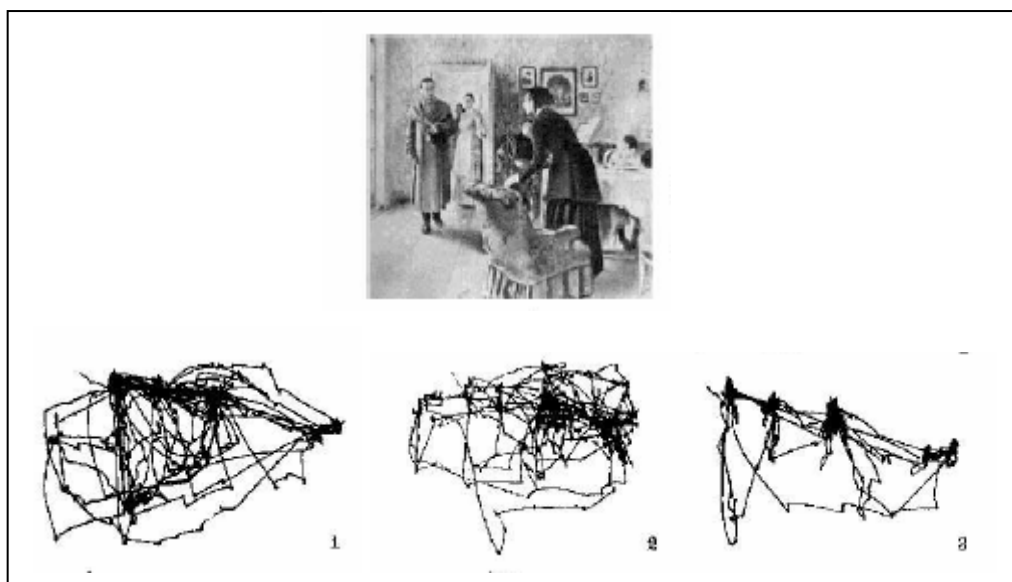


Figure 2 : Stimulus et patterns de mouvements oculaires mesurés selon le type d'instruction donnée aux participants dans l'étude de Yarbus (1967) : 1) Exploration libre de la peinture ; 2) Évaluation du confort matériel de la maisonnée ; 3) Estimation de l'âge de chacun des personnages (extrait de Yarbus, 1967)

Alors qu'il observe une relative constance dans les patterns oculaires de participants ayant pour consigne de regarder le tableau « Un visiteur inattendu » (Repin, 1884-1888), Yarbus (1967) observe que les patterns des mouvements oculaires changent drastiquement selon la question posée aux observateurs. Par exemple, s'il leur demande d'estimer l'âge de chacun des personnages, les sujets orientent leurs fixations oculaires sur les visages des protagonistes, alors que s'il leur demande d'évaluer le confort matériel de la maisonnée, ils orientent plutôt leurs fixations oculaires sur les meubles et les peintures accrochées au mur. Ces résultats, qui montrent que les participants extraient activement les informations de la scène requises par les demandes de la tâche, ont été maintes fois répliqués (Hayhoe, 2000; Hayhoe et al., 2002; Land & Hayhoe, 2001; Pelz, Hayhoe, & Loeber, 2001). Par exemple, lorsque les participants doivent préparer un sandwich ou une tasse de thé, leurs fixations oculaires sont systématiquement orientées vers les objets de la scène requis par la tâche (la théière, la tasse, le sucre...). Ainsi, sur les quelque 250 fixations enregistrées lors de la préparation d'un sandwich, Mary Hayhoe (2000) observe que seules une ou deux fixations oculaires concernent un élément de la scène non pertinent par rapport à la tâche (pour des résultats similaires, voir Land & Hayhoe, 2001). Plus précisément, les recherches couplant l'étude des patterns oculaires avec l'enregistrement des mouvements moteurs indiquent que lorsque des participants doivent réaliser une tâche (par exemple, préparer un sandwich), l'enchaînement des actions motrices (attraper le couteau, le plonger dans le pot de confiture...) reproduit pas à pas le pattern des fixations oculaires : chacun des éléments requis par la tâche momentanée est préalablement focalisé par l'œil (Hayhoe, Shrivastava, Mruczek, & Pelz, 2003; Land & Hayhoe, 2001). Les yeux constituent donc un outil privilégié pour

l'extraction progressive des informations exigées par la tâche (Hayhoe et al., 2002; Hayhoe et al., 2003; Land & Hayhoe, 2001; Triesch, Ballard, Hayhoe, & Sullivan, 2003). Ces exigences déterminent non seulement où doit être déployée l'attention, mais également quelle durée doit être allouée à l'extraction de telle ou telle information de la scène (Hayhoe et al., 2003), ce qui permet de réaliser avec succès des tâches complexes sans saturer les ressources cognitives (Ballard, Hayhoe, & Pelz, 1995; Ballard, Hayhoe, Pook, & Rao, 1998).

Plus généralement, tout élément de la scène pertinent par rapport à une tâche et stocké en mémoire de travail est susceptible d'attirer l'attention de façon endogène (Downing, 2000), et ce guidage de l'attention est tel qu'il est insensible aux objets de la scène non pertinents par rapport à la tâche, et ce, quelle que soit leur saillance perceptive. Ainsi, comme l'illustrent avec brio les recherches sur la vision sélective et la cécité inattentionnelle (Mack & Rock, 1998; Most et al., 2005; Simons, 2000a; Simons & Chabris, 1999), des éléments particulièrement saillants du point de vue perceptif peuvent être littéralement ignorés des traitements cognitifs au point de passer inaperçus des participants. Par exemple, dans une expérience où la tâche requiert de déterminer lequel de deux segments constitutifs d'une croix est le plus grand, les participants ne détectent pas l'apparition d'un stimulus critique non pertinent (un carré) dans leur champ visuel, même si celui-ci apparaît en position fovéale (Mack & Rock, 1998). Plus surprenant encore, des participants ayant reçu comme instruction de compter le nombre de passes échangées par les membres d'une même équipe sont dans l'incapacité de percevoir un élément extrêmement saillant des points du vue perceptif et contextuel (un gorille gesticulant), mais non pertinent, qui traverse la scène de part en part (Simons & Chabris, 1999). Pris ensemble, ces résultats constituent la parfaite illustration de l'influence déterminante des demandes de la tâche sur la perception d'une scène visuelle : l'œil et l'attention tendent à se concentrer sur les éléments pertinents par rapport à la tâche et à ignorer les autres éléments de la scène, quelle que puisse être leur saillance perceptive.

3. EN RÉSUMÉ...

Si les scènes visuelles naturelles sont très rapidement identifiées et appréhendées, elles restent sujettes aux interférences et nécessitent, pour résister aux interruptions saccadiques et au décours temporel, d'être consolidées en mémoire et requièrent, par conséquent, l'assistance des processus attentionnels (Intraub, 1981, 1999; Potter, 1999; Potter et al., 2002). L'attention visuelle joue donc un rôle capital dans la perception des scènes visuelles et des objets qu'elles contiennent, à un point tel qu'il n'est pas rare de considérer que l'activité de voir, ou du moins d'avoir conscience de ce que nous voyons, est indissociable des traitements attentionnels (Merikle & Joordens, 1997; Noë & O'Regan, 2000 ; Noë et al., 2000; Yantis, 2003). Dans ce chapitre, nous avons rappelé qu'il existait deux types d'informations visuelles susceptibles de guider l'orientation attentionnelle au sein des scènes visuelles et de permettre le traitement approfondi de tel ou tel aspect de la scène : les informations perceptivement saillantes (*eg.*, les apparitions brusques, le mouvement...), censées attirer le focus de l'attention de manière exogène, et les informations cognitivement saillantes (*eg.*, l'intérêt sémantique et la pertinence par rapport à la tâche), censées orienter le déploiement de l'attention, de manière endogène, vers les aspects de la scène qui méritent d'être traitées plus en détail. Dans ce chapitre, nous avons également montré qu'en dépit de leur extrême pertinence comportementale, il était délicat de considérer l'ascendant des processus exogènes sur les processus endogènes dans le guidage et le déploiement de l'orientation attentionnelle au sein des scènes visuelles. En effet, l'influence de facteurs endogènes, tels que la pertinence par rapport à la tâche, est à ce point influente qu'elle peut court-circuiter l'effet des facteurs exogènes sur l'orientation de l'attention, qui ont, pourtant, très probablement constitués les garants de la préservation de l'espèce humaine à travers les âges. Plus étonnant encore, l'influence de la tâche sur la distribution des ressources attentionnelles est telle que des

éléments extrêmement saillants des points de vue perceptifs et sémantiques peuvent passer inaperçus de participants engagés dans une tâche coûteuse en ressources cognitives, ce qui témoigne de la primauté accordée aux exigences de la tâche en matière de traitements visuels et attentionnels.

Le prochain chapitre propose quelques modèles de l'attention visuelle, qui permettent de rendre compte de l'influence des saillances perceptives et cognitives sur l'exploration de scènes visuelles plus ou moins complexes.

– CHAPITRE DEUX –

LES MODÈLES DE L'ORIENTATION ATTENTIONNELLE DANS LA PERCEPTION DE SCÈNES VISUELLES

Les modèles présentés dans ce chapitre visent à rendre compte de l'orientation attentionnelle au sein d'une scène et à estimer quels objets de la scène bénéficieront de traitements approfondis.

Lorsqu'il s'agit de réaliser une tâche de recherche visuelle dans un dispositif visuel simple dans lequel la cible (*eg.*, une barre) est un singleton défini par une orientation spécifique (*eg.*, 45° à droite), les déterminants du déploiement attentionnel semblent être relativement bien identifiés : les premières saccades semblent être uniquement dirigées par la saillance visuelle des items (c'est-à-dire vers tous les singletons d'orientation), et les saccades plus tardives par la saillance cognitive des items pertinents par rapport à la tâche (les items orientés à 45° vers la droite) (van Zoest, Donk, & Theeuwes, 2004). Par contre, la question du guidage attentionnel s'avère plus ardue lorsqu'il s'agit de scènes visuelles naturelles. En effet, dans des scènes visuelles naturelles, il est nécessaire, pour que l'attention visuelle soit efficacement dirigée vers les éléments de la scène que nous voulons voir plus en détail, qu'il existe au préalable des connaissances, ou, du moins, une représentation schématique de la scène qui puisse servir de cadre de référence aux déploiements attentionnels. Cette représentation devrait être opérationnelle dès les tous premiers instants de l'exploration visuelle de la scène, de manière à fournir une représentation approximative de la localisation des différents objets de la scène et permettre la focalisation et le traitement subséquents des

informations que nous voulons voir plus en détail. Très certainement, la saillance perceptive des éléments d'une scène naturelle n'a de cesse d'attirer, de manière exogène, l'attention et le regard. De récents travaux montrent d'ailleurs qu'il existe une corrélation positive entre la saillance perceptive des régions d'une scène et le nombre de fixations oculaires que celles-ci reçoivent (Chauvin, Herault, Marendaz, & Peyrin, 2002 ; Itti, 2005; Itti & Koch, 2000, 2001; Itti, Koch, & Niebur, 1998). Cependant, cette corrélation tend à diminuer lorsque la scène se dote d'une signification (Parkhurst, Law, & Niebur, 2002), ou quand une tâche doit être réalisée sur les scènes (Henderson & Hollingworth, 1999; Henderson et al., 1999; Land & Hayhoe, 2001), ce qui révèle que, à elle seule, la saillance perceptive des items ne permet pas de rendre compte du déploiement de l'attention visuelle dans la scène. En effet, lorsque des participants doivent réaliser une tâche dans une scène naturelle, l'étude des patterns de mouvements oculaires révèle que les premières saccades sur la scène sont très courtes (70 – 130 ms) et se posent indifféremment sur des objets visuels pertinents (48%) et non pertinents (52%) par rapport à la tâche à réaliser, alors que les saccades ultérieures sont plus longues (250 – 300 ms), et quasi exclusivement orientées vers les objets de la scène pertinents par rapport à la tâche, puisque les objets non pertinents ne sont plus la cible que de 16% des fixations oculaires (Hayhoe et al., 2003; Rao, Zelinsky, Hayhoe, & Ballard, 2002). Cette particularité des premières saccades oculaires est interprétée par Hayhoe et collaborateurs (2003) comme la signature d'un échantillonnage précoce des localisations approximatives des objets de la scène, préalablement à l'instanciation de la tâche. Cet échantillonnage serait destiné à fournir une représentation grossière de la position des objets de la scène, laquelle guiderait ensuite les yeux et l'attention vers la localisation des objets requis par la tâche, au fur et à mesure qu'ils deviendraient nécessaires (pour une conception similaire, voir Zelinsky, Rao, Hayhoe, & Ballard, 1997).

Par conséquent, les récents modèles de l'orientation de l'attention dans une scène visuelle ont intégré, en plus de la composante perceptive, une composante cognitive à l'orientation précoce de l'attention, en introduisant notamment le résumé, le schéma ou le contexte de la scène en tant que connaissances biaisant l'orientation attentionnelle purement basée sur la saillance visuelle des objets (Navalpakkam, Arbib, & Itti, 2005; Navalpakkam & Itti, 2002; Oliva, Torralba, Castelhana, & Henderson, 2003 ; Rensink, 2000a).

Ce chapitre présente quelques modèles de l'attention visuelle qui tentent de rendre compte de la perception des scènes visuelles en condition de vision active en intégrant une composante cognitive à la saillance perceptive dans le guidage de l'attention.

1. LES MODÈLES DE RÉFÉRENCE

1.1. Le modèle théorique de l'intégration de dimensions par l'attention

La théorie de l'intégration de dimensions par l'attention, proposée par Treisman et collaborateurs (Kahneman, Treisman, & Gibbs, 1992 ; Treisman, 1998; Treisman & Gelade, 1980), vise à rendre compte des performances observées dans des tâches de recherche visuelle d'une cible parmi un ensemble de distracteurs, et distingue deux étapes dans le traitement du dispositif visuel : un traitement préattentif, qui administre les caractéristiques de bas niveau de la scène, et un traitement attentif, qui permet de distinguer et d'isoler les objets de la scène (pour des approches théoriques similaires, voir Braun & Sagi, 1990; Irwin, 1992; Prinzmetal, Presti, & Posner, 1986)

Dans le modèle de Treisman et collaborateurs, la première étape du traitement visuel est dévolue au traitement des différents traits et dimensions¹⁷ basiques de la scène (couleurs, orientations, fréquences spatiales, mouvements...) qui sont précocement enregistrés et massivement traités en parallèle, de manière automatique, à travers le champ visuel (voir aussi Marr, 1982). Il résulte de ces traitements automatiques des cartes de dimensions qui codent grossièrement la localisation des traits de la scène les plus saillants. Cette représentation précoce des dimensions basiques de l'image permet la ségrégation de texture, le groupement différencié figure-fond, et la découverte rapide d'une cible définie par un trait ou une dimension spécifique (indépendamment de la taille du dispositif, Treisman & Gelade, 1980).

Par la suite, la perception des objets constitutifs de la scène requiert l'intervention de l'attention focalisée, qui agit comme un ciment intégrant toutes les dimensions basiques présentes dans le faisceau attentionnel en un objet unique. Ainsi, lorsque que l'issue des traitements préattentifs précoces ne permettait que de percevoir du rouge et des courbures, la focalisation de l'attention, permet, en unifiant les différents traits et dimensions de la scène, de percevoir une balle rouge. Une fois intégrés, les objets visuels sont stockés au sein de représentations épisodiques, les « fichiers d'objets¹⁸ », qui codent temporairement¹⁹ les descriptions et la forme des objets (Wolfe & Bennett, 1997). Ces « fichiers d'objets » peuvent ensuite être comparés aux connaissances conceptuelles stockées en mémoire à long terme, ce qui permet, a posteriori, la réussite des processus de reconnaissance et d'identification des

¹⁷ Dans ce modèle, les termes « dimension » et « trait » feront respectivement référence aux caractéristiques du stimulus qui sont prises en charge par des systèmes (ou canaux) perceptifs indépendants (*eg.*, la couleur, l'orientation, le contraste, le mouvement...) et aux valeurs particulières que peuvent prendre ces dimensions (*eg.*, rouge, vert...pour la dimension couleur ; horizontal, vertical...pour la dimension orientation).

¹⁸ En anglais « object files »

¹⁹ Récemment, il a été mis en évidence que les « fichiers d'objets » pouvaient exister pendant plusieurs secondes (Noles, Scholl, & Mitroff, 2005)

objets visuels. Enfin, une fois que l'attention se détourne d'eux, les objets tendent à se désintégrer jusqu'à retrouver l'état préattentif des traits et dimensions indifférenciés.

Dans ce modèle, le traitement d'un dispositif visuel se découpe donc en deux étapes. La première étape met en scène des processus préattentifs qui concourent à la création d'une représentation précoce et grossière des traits les plus saillants de la scène, et la seconde étape met en scène l'intervention de processus attentionnels soutenus, qui permettent de lier entre eux les différents traits présents dans le faisceau attentionnel, de sorte à permettre la perception d'un objet intégré.

Le modèle théorique proposé par Treisman et Gelade (1980) constitue, sans aucun doute, les fondations théoriques du modèle computationnel « Guided Search » (Wolfe, Cave, & Franzel, 1989) dont nous présenterons ici la version améliorée, « Guided Search 2 » (Wolfe, 1994)

1.2. Le modèle « Guided Search 2 »

Ce modèle est basé sur le modèle théorique de Treisman et Gelade (1980), avec lequel il partage de nombreuses caractéristiques. Élaboré dans le but de prédire et d'expliquer les performances humaines dans des tâches de recherche visuelle (dans lesquelles il s'agit de détecter la présence d'une cible définie par un ou plusieurs traits basiques), ce modèle distingue, tout comme le modèle de Treisman et Gelade (1980), deux niveaux dans le traitement d'un stimulus visuel : un niveau préattentif et un niveau attentif. Le modèle « Guided search 2 » se distingue néanmoins de son homologue théorique en attribuant un rôle important aux buts de la tâche dans le processus de recherche visuelle et en prenant avantage de la préparation attentionnelle aux caractéristiques visuelles de la cible.

Comme le montre la Figure 3, le modèle se décompose en trois étapes. Dans un premier temps, et tout comme le postulait déjà le modèle de Treisman et Gelade (1980), les différents traits et dimensions basiques du stimulus (couleur, orientation...) sont traités simultanément et de manière préattentive à travers le champ visuel, au moyen de canaux de traitement spécialisés pour chaque catégorie de traits basiques. Ces traitements parallèles conduisent à la création de plusieurs cartes de traits, qui codent chacune pour les régions du dispositif visuel les plus actives. Les régions les plus actives reçoivent par la suite le plus de traitements attentionnels. Dans ce modèle, les différents niveaux d'activation dépendent à la fois des processus de type bottom-up, c'est-à-dire de la saillance perceptive des traits du dispositif visuel, et des processus top-down, c'est-à-dire des buts de la tâche.

Le modèle calcule la saillance perceptive sur la base du degré de similarité perceptive que l'on peut observer entre un item et les items environnants, et ceci pour chaque trait basique : un item sera d'autant plus saillant et sa localisation dans l'espace d'autant plus active qu'il se distinguera des autres, et inversement, un item sera d'autant moins saillant et sa localisation dans l'espace d'autant moins active qu'il ressemblera aux autres (pour une appréhension de la saillance d'un stimulus sur la base de la similarité cible-distracteurs dans une tâche de recherche visuelle, voir aussi Duncan & Humphreys, 1989). La saillance cognitive est, quant à elle, définie par les caractéristiques qui font de la cible un élément unique du dispositif visuel. Ainsi, dans ce modèle, les traits qui individualisent la cible de la population des distracteurs sont fortement pondérés, et donc déterminent fortement l'orientation de l'attention (par exemple, si la cible se distingue des distracteurs parce qu'elle est le seul élément vert *et* rond, alors ces deux traits seront fortement pondérés par les processus top-down). Ainsi, chaque carte de trait code les régions du dispositif qui se révèlent être les plus actives du point de vue de la saillance perceptive *et* cognitive. Ces cartes de traits basiques parallèles et indépendantes sont ensuite fusionnées en une carte d'activation, qui

cumule les activations de chacune des cartes de traits, et la focalisation de l'attention est dirigée vers la région la plus active de la carte d'activation. Si la cible est trouvée (*ie.*, si la zone la plus active contient la cible), la tâche de recherche visuelle s'arrête ; sinon, le focus de l'attention est orienté vers la deuxième région la plus active de la carte d'activation, et la recherche de la cible se poursuit ainsi de suite par focalisation sérielle de l'attention vers les régions à niveau d'activation décroissant, jusqu'à découverte de la cible ou jusqu'à ce qu'un niveau minimal d'activation soit atteint, auquel cas la recherche visuelle, jugée infructueuse, est arrêtée (Chun & Wolfe, 1996).

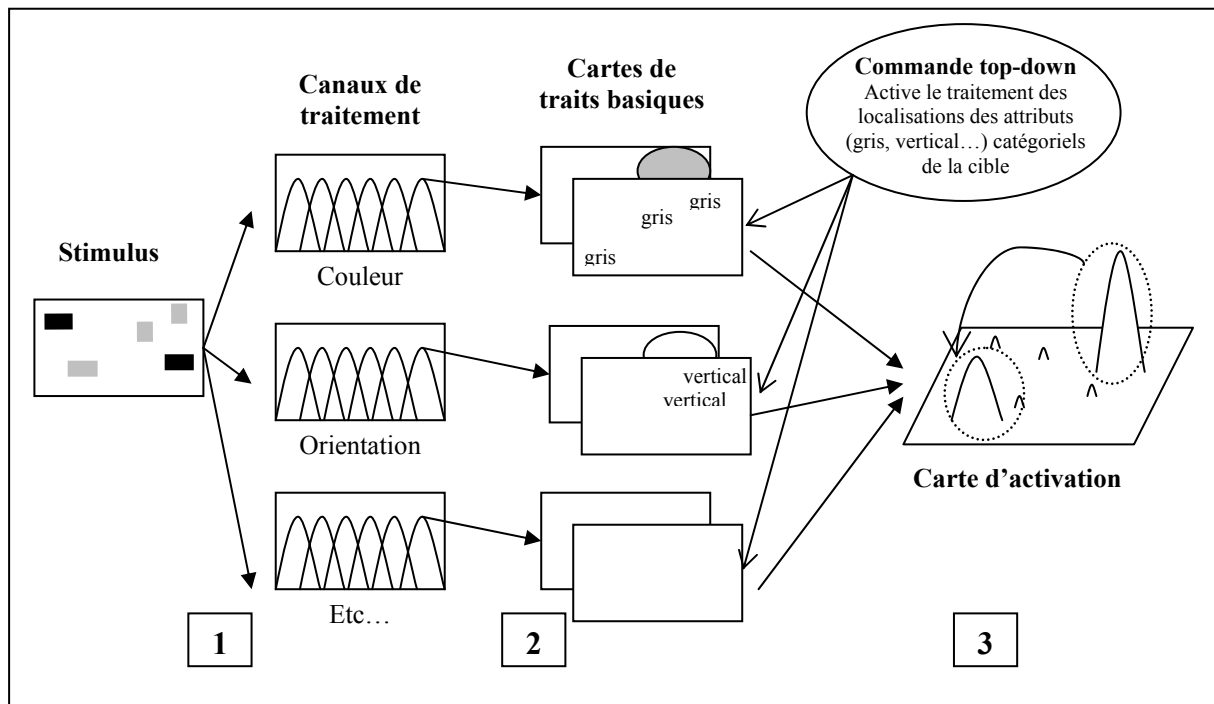


Figure 3 : Architecture du modèle « Guided Search 2 » (Wolfe, 1994). Ici, le modèle cherche une cible (un rectangle gris et vertical), définie par la conjonction des traits (gris et vertical) de deux dimensions basiques (couleur et orientation). 1) le stimulus est grossièrement traité en parallèle par des canaux de traitement spécifiques à chaque trait basique. 2) Ces traitements parallèles conduisent à l'émergence des cartes de traits basiques, dans lesquelles les zones les plus saillantes sont activées. La saillance résulte de l'action cumulée des différences visuelles locales (traitement de type bottom-up) et des demandes de la tâche (top-down). 3) Les différentes activations sont ensuite combinées au sein d'une carte d'activation, qui va guider l'attention focalisée successivement vers les zones les plus actives, dans un ordre décroissant, jusqu'à découverte de la cible ou abandon de la recherche visuelle.

1.3. Avantages et lacunes de ces modèles

Un des principaux points forts des modèles proposés par Treisman & Gelade (1980) et Wolfe (1994) est de postuler que le traitement visuel se découpe en deux étapes, préattentive et attentive, découpage dont la plausibilité biologique est assurée : il existe en effet des canaux spécialisés dans le traitement des différents traits et dimensions de la scène (voir Bullier, 1998), et le résultat de ces traitements peut tout à fait servir de guide à l'orientation exogène de l'attention (Wolfe & Horowitz, 2004). De plus, ces deux modèles présentent l'avantage de pouvoir rendre compte des performances humaines observées dans des tâches de recherches visuelles simples (dans lesquelles l'efficacité de la recherche ne dépend pas de la taille du dispositif) et plus complexes (comme c'est le cas des recherches de conjonctions de traits où l'efficacité de la recherche dépend de la taille du dispositif et du degré de similarité que les distracteurs partagent avec la cible), et la notion de « fichier d'objet » postulée par Treisman (Kahneman et al., 1992; Treisman, 1998; Treisman & Gelade, 1980) est cohérente avec les limitations des ressources attentionnelles et mnésiques (Irwin, 1992 ; Luck & Vogel, 1997; Rensink, 2000c; Vogel, Woodman, & Luck, 2001). Enfin, en proposant que les buts de la tâche biaisent directement les cartes de traits basiques, le modèle proposé par Wolfe (1994) évite d'avoir à postuler une carte de saillance qui, selon certains auteurs, est plus théorique que biologique (Desimone & Duncan, 1995).

Toutefois, un des inconvénients majeurs de ces modèles est qu'ils rendent difficilement compte du traitement des scènes visuelles naturelles, qui constituent un matériel bien plus complexe que les dispositifs visuels utilisés en laboratoire. En effet, il est difficile pour les modèles de prédire avec succès dans quelle mesure l'attention sera orientée vers telle ou telle région lorsque une scène visuelle complexe est présentée, puisque aucun des modèles proposés par Treisman & Gelade (1980) et Wolfe (1994) n'intègre les connaissances sur la

scène stockées en mémoire à long terme et supposées servir de guide aux déplacements attentionnels dans la scène, à savoir le résumé, le contexte et le schéma de scène.

De plus, le modèle « Guided search 2 » présente l'inconvénient de nécessiter que les propriétés visuelles de la cible soient connues *préalablement* à l'initiation de la recherche visuelle, puisqu'il postule que les cartes de traits sont biaisées, très précocement, par les caractéristiques visuelles de la cible dont on sait qu'elles distinguent la cible des distracteurs. Or, il n'est pas si fréquent, dans la vie de tous les jours, de rechercher une cible dont on connaît parfaitement les caractéristiques visuelles : par exemple, si je dois retrouver un numéro de téléphone gribouillé à la hâte sur une des pages de mon calepin, je ne peux réellement orienter stratégiquement ma tâche de recherche visuelle ni sur les caractéristiques visuelles du support papier (toutes les pages de mon calepin se ressemblent), ni sur les caractéristiques visuelles du numéro que j'y ai inscrit (contient-il un zéro ? un deux ?). Dans ce genre de tâches de recherche visuelle, rarissimes en laboratoire et relativement fréquentes dans la vie de tous les jours, les prédictions du modèle proposé par Wolfe (1994) risquent d'être hasardeuses.

2. LES MODÈLES DE LA PERCEPTION DE SCÈNES NATURELLES

Les modèles de l'attention visuelle proposés récemment visent à rendre compte de l'allocation de l'attention focalisée lorsque nous devons réaliser une tâche de recherche visuelle dans des scènes naturelles. Cette section ne présentera que les modèles qui intègrent des connaissances sur la scène (contexte, résumé, schéma) dans le déploiement de l'attention au sein des scènes visuelles (et cela bien que nombreux soient les modèles qui conçoivent le déploiement attentionnel sur la base de la saillance perceptive des régions de la scène:

Chauvin et al., 2002; Itti, 2005; Itti & Koch, 2000; Itti et al., 1998; Parkhurst et al., 2002). Ce choix résulte du fait que ces connaissances jouent très certainement un rôle dans le guidage de l'attention au sein de la scène (Biederman, 1972; De Graef, 1992; Friedman, 1979; Intraub, 1997; Oliva, 2005; Torralba, 2003) et que la corrélation observée entre la densité des fixations sur tel ou tel aspect de la scène et les régions estimées les plus perceptivement saillantes par les modèles tend à s'amenuiser quand la scène présentée est dotée de sens (Parkhurst et al., 2002; Parkhurst & Niebur, 2004).

2.1. Architecture triadique de la vision

L'architecture triadique de la vision proposée par Rensink (Rensink, 2000a, 2000b, 2001, 2002; Rensink, O'Regan, & Clark, 2000) postule que trois systèmes indépendants permettent la perception des scènes visuelles (voir Figure 4). Un premier système, correspondant à la vision de bas niveau, opère dès les premières étapes du traitement visuel de la scène et conduit à l'émergence de structures visuelles, hautement détaillées et extrêmement volatiles, qui disparaissent aussitôt que la lumière cesse de pénétrer dans l'œil. Dans ce modèle, le second niveau de la perception de scènes visuelles fait référence à un système attentionnel à capacité limitée, qui transforme ces structures labiles en objets intégrés, représentations stables et cohérentes dans le temps et l'espace. Le troisième niveau fait état d'un système non attentionnel, à capacité limitée, qui fournit un cadre de référence au déploiement attentionnel.

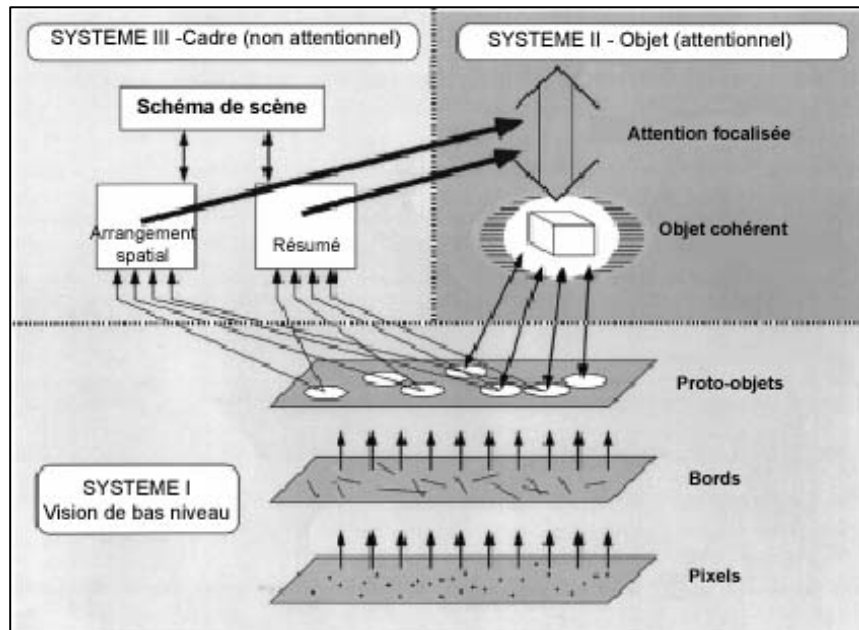


Figure 4 : Architecture triadique de la vision selon Rensink (2000). Le modèle postule que la perception des scènes visuelles repose sur l'interaction de trois systèmes. Le premier système correspondant à la vision de bas niveau, le deuxième système (attentionnel) qui permet la perception des objets, et le troisième système (non attentionnel) sert de cadre de référence au déploiement de l'attention focalisée dans la scène (extrait et traduit de Rensink, 2000a)

Les premières étapes du traitement d'une scène visuelle sont, dans le modèle de Rensink, similaires au système préattentif postulé par Treisman et Gelade (1980). Rensink propose en effet que, dès présentation d'une scène visuelle, les dimensions et traits basiques de la scène sont massivement traités de manière parallèle et préattentive, et que le résultat de ces traitements fournit, en quelques centaines de millisecondes, une représentation rétinotopique de la scène (voir aussi Marr, 1982). Cette représentation, hautement détaillée, serait extrêmement volatile puisqu'elle n'existerait que le temps où la lumière pénétrerait dans l'œil et ne survivrait donc ni aux clignements de paupières ni aux saccades oculaires (Rensink, O'Regan, & Clark, 1997; Rensink et al., 2000). Enfin, cette représentation serait composée de proto-objets (ou objets préattentifs) (Wolfe, 1999; Wolfe & Bennett, 1997), assemblages complexes de traits et de fragments correspondant à des structures localisées dans la scène, qui constitueraient les structures de plus bas niveau accessibles aux traitements

attentionnels, sans lesquels il serait impossible d'accéder à la perception, cohérente des points de vue spatial et temporel, des objets de la scène.

Le deuxième système postulé par Rensink est un système attentionnel à capacité limitée, par l'action duquel les caractéristiques visuelles portées par les proto-objets sont liées entre elles afin de former des objets visuels unifiés. Dans ce modèle, en effet, les proto-objets n'acquièrent la cohérence indispensable à leur survie spatio-temporelle que sous le joug de l'attention focalisée, qui organise, traite et lie entre elles ces structures précoces, instables et volatiles en objets intégrés, ce qui leur permet de conserver identité et continuité dans le temps et dans l'espace (Wolfe, 1999, Rensink, 2000). Ainsi, seule la focalisation de l'attention sur ces structures serait en mesure de leur conférer la cohérence spatiotemporelle indispensable à leur conversion en objet stable, et à leur survie identitaire. Dans ce modèle, l'attention focalisée serait donc, tout comme le postulaient déjà Treisman et Gelade (1980), intimement liée à la perception des objets. Cependant, à la différence des postulats de Treisman qui autorisent la coexistence de plusieurs « fichiers d'objets », l'effet intégratif de l'attention focalisée se limiterait, dans le modèle de Rensink, à un objet unique représenté au sein d'un nexus (structure stable et unique contenant une description sommaire de l'objet focalisé par l'attention, et qui sous-tend la cohérence spatiotemporelle), chaque objet ne pouvant, de plus, n'être représenté que par 4 à 6 de ses propriétés constitutives, conformément aux limitations attentionnelles (Irwin, 1992; Luck & Vogel, 1997; Pashler, 1988; Vogel et al., 2001). Une fois l'attention détournée de lui, l'objet intégré perdrait sa cohérence spatiotemporelle et retrouverait l'état labile des proto-objets, sans qu'il y ait de post-effet de l'attention (ce point est par ailleurs agréé par Wolfe, 1999), si ce n'est une trace mnésique plus conceptuelle que visuelle, stockée en mémoire à court terme.

Enfin, dans le modèle, ces deux systèmes de représentation (système de bas niveau et système attentionnel à capacité limitée) s'accompagnent d'un troisième système, non

attentionnel, activé par les informations visuelles véhiculées par les proto-objets et qui fournit un schéma global de la scène, lequel sert de guide aux déplacements de l'attention. Ce schéma global résulte de l'action conjuguée des connaissances précocement activées par la présentation de la scène (le résumé et l'arrangement spatial de la scène) et des connaissances conceptuelles stockées en mémoire à long terme (le schéma de scène), qui, ensemble, permettent de guider efficacement l'attention vers les objets de la scène que nous voulons voir plus en détail, afin qu'ils soient tour à tour focalisés par l'attention et intégrés dans un nexus.

2.2. Un modèle computationnel de l'attention visuelle

Navalpakkam et Itti (Navalpakkam et al., 2005; Navalpakkam & Itti, 2002, 2005) reprennent le principe de l'architecture triadique de Rensink, et proposent un modèle biologiquement plausible de l'orientation attentionnelle au sein d'une scène visuelle, modèle qui prend en compte à la fois la saillance perceptive et la saillance cognitive des informations visuelles véhiculées par la scène. Ce modèle est construit sur les principes qui gouvernent les modèles du guidage bottom-up de l'attention visuelle (Itti, 2005; Itti & Koch, 2000, 2001; Itti et al., 1998; Parkhurst et al., 2002), dans lesquels plusieurs cartes de traits et dimensions (couleur, orientation, intensité...) sont très précocement construites en parallèle à travers le champ visuel, et combinées en une carte de saillance. L'attention est ensuite déployée de façon sérielle vers les zones les plus actives de la carte de saillance, dans un ordre décroissant.

La supériorité du modèle computationnel de l'attention visuelle proposé par Navalpakkam et Itti (2005) sur les modèles du guidage bottom-up de l'attention visuelle est de prendre en compte l'influence des demandes de la tâche sur le déploiement et l'orientation de l'attention focalisée dans une scène visuelle complexe. Illustré en Figure 5, il se décompose en quatre phases.

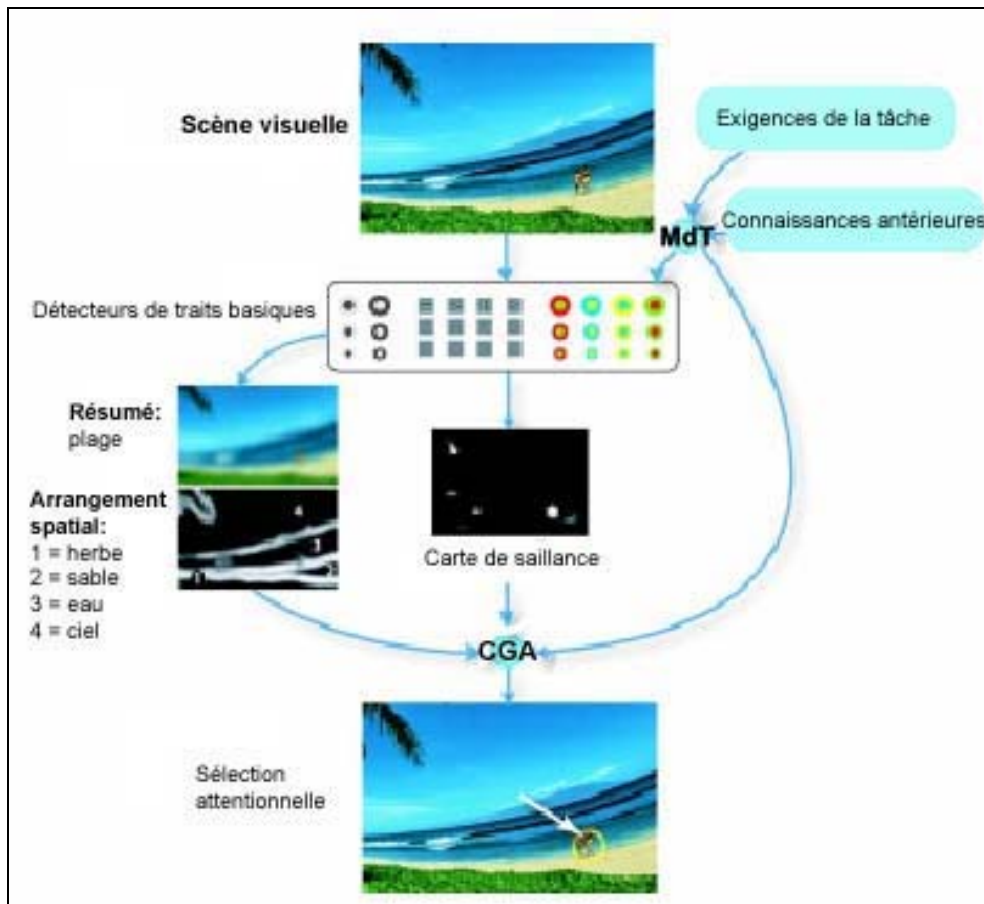


Figure 5. Architecture théorique du modèle computationnel proposé par Navalpakkam et Itti (Navalpakkam et al., 2005; Navalpakkam & Itti, 2002, 2005) ; MdT = Mémoire de Travail ; CGA = Carte de Guidage de l'Attention

La première phase du modèle commence alors même que la scène n'est pas encore présentée, ce que les auteurs nomment «phase yeux fermés ». Dans cette phase, la mémoire de travail est initialisée à partir des mots clés qui définissent la tâche. Par exemple, si la tâche est de localiser les êtres humains présents dans la scène, la mémoire de travail va être formatée, sur la base des exigences de la tâche, afin de distinguer les objets pertinents (ici, les êtres humains) des objets non pertinents par rapport à la tâche de recherche visuelle. Le modèle considère que, dans cette première étape, l'initialisation de la mémoire de travail peut également tirer parti des connaissances stockées en mémoire à long terme (ce qui s'apparenterait au schéma de scène) pour définir les caractéristiques de la cible, et se préparer à répondre à tel ou tel trait caractéristique de la cible.

La deuxième phase du modèle est une phase de « computation », qui opère dès que la scène est présentée. Dans cette phase, les processus visuels de bas niveau, biaisés pour répondre de façon appuyée aux traits de la cible, traitent massivement les traits et dimensions basiques de la scène de manière préattentive. Les résultats de ces traitements sont ensuite combinés en une carte de saillance, qui est une représentation topographique de la scène en deux dimensions codant la saillance de toutes les localisations de la scène (lesquelles sont, rappelons-le, biaisées en fonction des caractéristiques de la cible). Les régions les plus saillantes de cette représentation sont les plus susceptibles d'attirer l'attention. La carte de saillance est ensuite combinée aux connaissances activées très précocement dès la présentation de la scène (ici, le résumé de la scène et son arrangement spatial) afin de fournir une carte de guidage de l'attention, qui intègre à la fois les exigences de la tâche, la saillance perceptive des différentes régions de la scène, et les connaissances sur la scène. L'attention focalisée est déployée vers les régions les plus actives de la carte de guidage de l'attention, dans l'ordre décroissant de leur niveau d'activation.

Les troisième et quatrième phases du modèle rendent compte des traitements attentionnels : les traits de la région de la scène focalisée par l'attention sont intégrés en une entité visuelle dont la mémoire de travail estime la pertinence par rapport à la tâche et à laquelle elle attribue une cote de pertinence. Si l'entité extraite de la scène est peu, voire non pertinente par rapport à la tâche, elle est cotée d'un niveau d'activation inférieur à 1.00, ce qui en fait une entité à ignorer lors des fixations attentionnelles ultérieures. Si, par contre, l'entité extraite de la scène est pertinente par rapport à la tâche, elle est cotée d'un niveau d'activation supérieur à 1.00, ce qui en fait une entité d'intérêt vers laquelle il faudra orienter de nouvelles fixations attentionnelles (on retrouve un postulat similaire dans Henderson et al., 1999). Par la suite, l'orientation du faisceau attentionnel se stabilise et se concentre, de manière récurrente, sur les régions désignées d'intérêt de la scène.

2.3. Avantages et inconvénients des modèles présentés

Une des qualités fondamentales des modèles proposés par Rensink (2000) et Navalpakkam et Itti (2005) est de prendre en compte les connaissances activées dès la présentation d'une scène visuelle, et dont on suppose qu'elles participent activement au déploiement du focus de l'attention au sein des scènes visuelles naturelles (De Graef, 1992; Henderson & Hollingworth, 1999; Introub, 1981; Oliva, 2005; Potter, 1976, 1993, 1999; Potter et al., 2004; Potter et al., 2002; Sanocki, 2003), ce que ne faisaient ni le modèle de recherche visuelle (Kahneman et al., 1992; Treisman, 1998; Treisman & Gelade, 1980; Wolfe, 1994; Wolfe et al., 1989), ni les modèles computationnels de l'orientation bottom-up de l'attention (Chauvin et al., 2002 ; Itti & Koch, 2000; Itti et al., 1998; Parkhurst et al., 2002; Parkhurst & Niebur, 2004). Bien entendu, d'autres modèles de la perception de scènes visuelles complexes, qui ne sont pas présentés ici, ont pris également le parti de modéliser l'influence des connaissances sur la scène dans le guidage de l'attention (Oliva et al., 2003; Torralba, 2003). Ces modèles fonctionnent sur le même principe que le modèle computationnel de Navalpakkam et Itti (2005), à savoir que la carte de saillance est biaisée par les connaissances sur la scène, de sorte que le guidage de l'attention résulte de l'action combinée de la saillance perceptive et cognitive. Ainsi, Torralba (2003) et Oliva et collaborateurs (2003) proposent que le guidage de l'attention dépend des informations véhiculées par le contexte dans lesquels les objets sont enchâssés, et basent, dans leurs modèles, le guidage du focus attentionnel sur les régularités statistiques observées entre la distribution des traits basiques de la scène et la localisation d'objets spécifiques (voir Oliva & Schyns, 2000; Schyns & Oliva, 1994; Torralba & Oliva, 2003). L'originalité de ces modèles est de tirer avantage de la covariation entre les traits basiques de la scène et les antécédences

contextuelles afin de guider l'attention vers les régions d'intérêt de la scène, susceptibles de contenir la cible.

Tous ces modèles présentent donc l'avantage certain de considérer à la fois les effets de la saillance visuelle, des connaissances sur la scène et de la saillance cognitive (*ie.*, les exigences de la tâche) des informations visuelles dans l'orientation du focus de l'attention vers telle ou telle région de la scène. En plus d'être biologiquement plausibles, ces modèles computationnels présentent, de surcroît, l'avantage de permettre la formulation d'hypothèses précises et immédiatement testables, ainsi que de comparer les prédictions des modèles aux mouvements oculaires enregistrés chez des sujets humains.

Par contraste, le modèle de Rensink (2000) reste plus que vague au sujet de l'influence de la tâche sur le guidage de l'attention. En effet, ce modèle postule que les connaissances sur la scène (le résumé, l'arrangement spatial et le schéma de scène) vont guider le focus de l'attention vers les zones d'intérêt de la scène. Ainsi, en condition de vision passive (c'est-à-dire lorsque les participants reçoivent l'instruction d'explorer librement la scène), il est aisé d'exprimer des prédictions solides à partir des postulats de ce modèle : les zones d'intérêt sémantique de la scène bénéficieront de la majorité des traitements attentionnels. Par contre, il est difficile pour le modèle d'émettre des prédictions fiables en condition de vision active (c'est-à-dire lorsque les participants doivent réaliser une tâche sur les scènes visuelles), puisqu'il ne prend pas réellement en compte les exigences de la tâche comme facteur du déploiement attentionnel. De plus, en n'autorisant que la représentation d'objets uniques et en postulant par ailleurs l'absence de post-effets attentionnels sur la représentation d'un élément en mémoire, le modèle peut difficilement rendre compte des performances obtenues dans certaines expériences, où les participants sont en mesure de signaler qu'un changement est survenu sur un objet alors même que l'attention s'en est dégagée (Hollingworth & Henderson, 2002).

Le modèle computationnel, proposé par Navalpakkam et collaborateurs (2005), n'est pas non plus exempt de zones d'ombres. Par exemple, il nécessite, comme le modèle « Guided search 2 » de Wolfe (1994), que soit connues au préalable les caractéristiques visuelles de la cible, ce qui en diminue nécessairement l'efficacité lorsque les caractéristiques de la cible sont méconnues. De plus, le modèle fait l'hypothèse d'une carte de saillance, dont les modalités de création restent discutées. En effet, soit on suppose que la carte de saillance est créée une fois pour toutes pour l'ensemble de la scène durant la première fixation, soit on suppose qu'elle est créée à chaque nouvelle fixation oculaire. Ces deux alternatives s'avèrent pour l'instant incertaines et problématiques : en effet, la première hypothèse nécessite que cette représentation résiste aux saccades oculaires, ce qui serait en parfait désaccord avec les recherches ayant montré que l'information métrique précise n'était pas retenue pendant une saccade oculaire (Bridgeman, van der Heijden, & Velichkovsky, 1994; Liversedge & Findley, 2000; McConkie & Currie, 1996). La seconde hypothèse nécessite, quant à elle, que ce soit le processus d'inhibition de retour²⁰ qui soit retenu pendant les saccades oculaires, de façon à ce que l'attention ne se focalise pas sur une même localisation à chaque fixation²¹. Enfin, se pose toujours le problème de la localisation des cartes de saillance et de guidage de l'attention dans le cerveau, bien que de récentes études suggèrent que de telles cartes pourraient être prises en charge par des aires contenues dans le cortex pariétal postérieur (Assad, 2003; Colby & Goldberg, 1999 ; Corbetta & Shulman, 2002; Gottlieb, Kusunoki, & Goldberg, 2005 ; Kusunoki, Gottlieb, & Goldberg, 2000; Mazer & Gallant, 2003 ; Treue, 2003), lesquelles contiennent des neurones qui déchargent à la fois lorsqu'un stimulus pénètre soudainement

²⁰ Le processus d'inhibition de retour, qui empêche l'attention de focaliser obstinément sur la zone la plus active de la carte de saillance, est d'ailleurs implémenté dans une bonne partie des modèles computationnels présentés

²¹ Cette dernière hypothèse est, peut-être, plus réaliste que la première

dans leur champ récepteur (saillance visuelle) et lorsqu'un stimulus, déjà présent dans le champ récepteur des neurones, devient pertinent par rapport à une tâche (saillance cognitive).

3. ATTENTION... REPRÉSENTATION ?

Tous les modèles présentés considèrent les influences conjuguées des saillances visuelles et cognitives de la scène et des connaissances stockées en mémoire, dans le but de rendre compte du déploiement de l'attention visuelle au sein des scènes plus ou moins complexes. Aussi, étant donné que la focalisation attentionnelle est un évènement indispensable au traitement plus en détail et à la représentation en mémoire des informations visuelles (Henderson & Hollingworth, 2003; Noë & O'Regan, 2000; O'Regan, 2001; Rensink, 2001; 2002 ; Rensink et al., 1997 ; Simons, Mitroff, & Franconeri, 2003), ces modèles devraient donc nous permettre d'entrevoir quelles informations portées par les scènes visuelles seront représentées en mémoire. Or, dans leur version computationnelle (Navalpakkam et al., 2005; Navalpakkam & Itti, 2002, 2005; Oliva et al., 2003; Torralba, 2003; Wolfe, 1994), ces modèles appréhendent le focus de l'attention du point de vue des fixations oculaires, de sorte que la similarité observée entre les performances des modèles et les performances de sujets humains, interprétée comme la signature de la capacité des modèles à simuler les comportements attentionnels humains, signe, en fait, la capacité des modèles à simuler les comportements *oculaires* humains. Ce faisant, ces modèles sont incapables de rendre compte des informations de la scène qui seront représentées en mémoire puisqu'il est aujourd'hui communément admis qu'il existe un goulot d'étranglement entre la vision et la mémoire (Averbach & Coriell, 1961; Pashler, 1988; Phillips, 1974), et que l'on estime la capacité de stockage de la mémoire de travail visuelle (ou mémoire visuelle à court

terme) à 4 – 5 objets complexes²² (Irwin, 1992; Lee & Chun, 2001; Luck & Vogel, 1997; Vogel et al., 2001). Il est donc impossible que tous les éléments de la scène fixés par l'œil bénéficient d'un codage ultérieur en mémoire. Par ailleurs, si l'on admet, comme le considèrent Liversedge et Findley (2000), que les mouvements oculaires sont un bon indicateur de l'allocation attentionnelle, il reste tout de même délicat de formuler une estimation précise concernant les objets de la scène qui seront représentés en mémoire puisque rien ne garantit que les objets préalablement focalisés par l'attention soient représentés en mémoire (Horowitz & Wolfe, 1998; 2001 ; Simons & Levin, 1998), notamment en raison de la compétition qui s'exerce entre les différents objets visuels pour s'assurer d'une représentation ultérieure et d'un accès conscient en mémoire (Desimone & Duncan, 1995; Reynolds, Chelazzi, & Desimone, 1999 ; VanRullen & Koch, 2003). Pourtant, la question de la représentation en mémoire est cruciale, puisque ce sont ces mêmes informations représentées en mémoire qui seront les plus susceptibles de servir nos comportements ultérieurs (Desimone & Duncan, 1995 ; Ward & Duncan, 1996). Aussi, le chapitre suivant s'intéresse plus précisément à la question de la représentation en mémoire des scènes visuelles complexes et des informations qu'elles véhiculent.

²² Ou objets intégrés

– CHAPITRE TROIS –

LA REPRÉSENTATION EN MÉMOIRE DES SCÈNES VISUELLES COMPLEXES

La question de la représentation en mémoire des scènes visuelles complexes a rencontré un regain d'intérêt avec la résurgence du phénomène de cécité au changement (Di Lollo, 1980; McConkie & Zola, 1979; O'Regan, 2001; Pashler, 1988; Phillips, 1974; Rensink, 2001, 2002; Simons, 2000b; Simons & Ambinder, 2005; Simons & Levin, 1997; Simons & Rensink, 2005). La cécité au changement fait référence à la surprenante difficulté de détecter (McConkie & Currie, 1996; Phillips, 1974; Scott-Brown, Baker, & Orbach, 2000), de localiser (Fernandez-Duque & Thornton, 2000; Smilek, Eastwood, & Merikle, 2000) ou d'identifier (Mondy & Coltheart, 2000; Watanabee, 2003) un changement opéré sur une scène pendant une interruption visuelle. En effet, lorsque le changement est concomitant avec une interruption visuelle, le réflexe de fixation, inéluctablement engendré par la modification opérée dans la scène, est court-circuité : l'attention n'est plus orientée vers la localisation du changement sur la base de ce facteur exogène, et la réussite de la détection des changements ne peut plus s'appuyer que sur des processus mnésiques pour déterminer quel item de la scène a été changé²³. Aussi, ces dernières années ont vu la cécité au changement être utilisée comme

²³ Lorsqu'au contraire, le changement n'est pas concomitant avec une interruption visuelle, l'attention est automatiquement dirigée vers la région du changement, lequel est immédiatement détecté (Blackmore et al., 1995 ; Rensink et al., 1997)

un moyen d'étudier la représentation en mémoire des objets et des scènes visuelles (Simons & Levin, 1997 ; Simons et al., 2003).

La cécité au changement est un phénomène robuste, qui se manifeste dès que le changement réalisé sur la scène est masqué par une interruption visuelle, que cette interruption soit le fait d'une saccade oculaire (Carlson-Radvansky & Irwin, 1995; Irwin, 1992; McConkie & Currie, 1996; McConkie & Zola, 1979), d'un clignement de paupières (O'Regan et al., 2000), ou qu'elle résulte au contraire d'une intervention expérimentale comme un décalage de l'image par rapport à sa position d'origine (Blackmore et al., 1995), l'insertion d'un masque local²⁴ (O'Regan, 2001; O'Regan, Rensink, & Clark, 1999) ou global²⁵ (Pashler, 1988; Phillips, 1974; Rensink et al., 2000; Simons, 1996), une coupure dans la dynamique de la scène²⁶ (Levin & Simons, 1997), ou même quand il n'y a pas, à proprement parler, d'interruption visuelle, à condition que la survenue du changement se fasse graduellement, de sorte à ne pas attirer l'attention de manière exogène (Simons, Franconeri, & Reimer, 2000). La cécité au changement se retrouve quels que soient les changements effectués sur la scène (addition, délétion, changement de couleur, déplacement, ou rotation d'un objet de la scène, Aginsky & Tarr, 2000; Blackmore et al., 1995; Cole, Kentridge, Gellatly, & Heywood, 2003 ; Mondy & Coltheart, 2000; O'Regan et al., 1999), et se manifeste autant sur du matériel visuel simple (figures géométriques, dispositif de lettres ou de chiffres..., Becker & Pashler, 2002; Di Lollo, 1980; Fernandez-Duque & Thornton, 2000; Landman, Spekreijse, & Lamme, 2003; Phillips, 1974; Scott-Brown et al., 2000 ; Stolz & Jolicoeur, 2004 ; Thornton & Fernandez-Duque, 2000; Tijus & Reeves, 2004) que plus complexe, qu'il soit statique (dessins de scènes naturelles ou de formes complexes et

²⁴ Qui recouvre quelques zones de la scène seulement

²⁵ Qui recouvre l'intégralité de la scène

²⁶ Cette méthode implique le même procédé de changement de plan que celui utilisé au cinéma

ambiguës, photographies de scènes naturelles...Blackmore et al., 1995; McConkie & Currie, 1996; Mondy & Coltheart, 2000; O'Regan et al., 1999; Simons & Ambinder, 2005; Simons et al., 2000; Tatler, Gilchrist, & Rusted, 2003; Werner & Bjorn, 2000; Williams & Simons, 2000), ou dynamique (Levin & Simons, 1997; Shinoda, Hayhoe, & Shrivastava, 2001; Wallis & Bülthoff, 2000), voire plus écologique encore dans le cas des cécités au changement observées dans des interactions avec le monde réel (Angelone, Levin, & Simons, 2003; Levin, Simons, Angelone, & Chabris, 2002; Simons, Chabris, Schnur, & Levin, 2002; Simons & Levin, 1998). Enfin, le phénomène de cécité au changement ne semble dépendre ni des durées de présentation du stimulus (Rensink et al., 2000), ni de la durée ou de l'apparence du masque visuel utilisé (Rensink et al., 2000 ; Shore & Klein, 2000; Tijus & Reeves, 2004). Enfin, il se retrouve quel que soit le type de réponses recueilli, qu'il s'agisse de réponses explicites (report verbal du changement détecté, latences de détections..., Mitroff & Simons, 2002; Shinoda et al., 2001; Simons, 1996), ou implicites (saccades oculaires en direction du changement perçu, rapidité de la réponse à une autre tâche..., Fernandez-Duque & Thornton, 2000; Hayhoe, 2000; Mitroff, Simons, & Franconeri, 2002; Thornton & Fernandez-Duque, 2000).

Toutes les études sur la cécité aux changements s'entendent sur un point : l'attention focalisée sur un objet est requise pour que son changement puisse être détecté (Hollingworth, Williams, & Henderson, 2001; Levin, Simons et al., 2002; Mondy & Coltheart, 2000; Noë & O'Regan, 2000; Noë et al., 2000; O'Regan, 2001; O'Regan et al., 2000; O'Regan et al., 1999; Rensink, 2000c, 2000d, 2002; Rensink et al., 1997, 2000; Scholl, 2000; Simons, 2000b; Simons & Ambinder, 2005; Simons & Levin, 1997; Williams & Simons, 2000), mais bien que nécessaire, celle-ci n'est pas un facteur suffisant à la représentation en mémoire des informations visuelles. Ainsi, la détection des changements n'est pas systématiquement assurée lorsque l'objet du changement est focalisé par l'œil et/ou l'attention, comme en

témoignent plusieurs recherches (Levin, Simons et al., 2002; O'Regan et al., 2000; Simons & Levin, 1998). Par exemple, dans la recherche menée par O'Regan et collaborateurs, 40% des changements survenant sur des objets visuels pourtant focalisés par l'œil ne sont pas détectés par les participants, et les recherches sur la cécité au changement menées dans le cadre d'interactions avec le monde réel²⁷ montrent que près de la moitié des participants ne s'aperçoit pas qu'une autre personne s'est substituée à celle avec laquelle ils s'entretenaient alors même qu'elle est la cible de leur attention.

Par conséquent, le parti pris des expériences sur la cécité au changement a été le suivant : puisque les changements ne s'accompagnent pas de facteurs exogènes susceptibles d'accaparer les traitements attentionnels et de favoriser leur détection, alors on peut considérer que les changements qui ont été détectés par les participants témoignent des informations qui ont été focalisées par l'attention et représentées en mémoire de façon suffisamment précise pour soutenir la détection de leur changement. Ainsi, les facteurs de la détection des changements sont le reflet des facteurs de la représentation en mémoire des informations visuelles (Simons & Ambinder, 2005; Simons & Rensink, 2005). Aussi, les sections suivantes proposent une synthèse des facteurs influençant la représentation des informations contenues dans les scènes visuelles en mémoire.

²⁷ Classiquement, dans ces expériences, l'expérimentateur va à la rencontre d'un piéton et lui demande le chemin à suivre pour se rendre à un point X. Pendant que le sujet donne ses indications, le contact visuel entre l'expérimentateur et le sujet est brièvement interrompu par le passage de deux compères transportant une vitre opaque. Pendant cette interruption visuelle, l'expérimentateur est remplacé par l'un des compères. Alors que les deux personnes interchangeables diffèrent par leur physionomie, leurs vêtements, leur couleur de cheveux et leur voix, rares sont les participants qui rapportent, a posteriori, un changement d'interlocuteur, même lorsqu'ils sont directement questionnés sur ce point (Levin, Simons et al., 2002; Simons et al., 2002; Simons & Levin, 1998).

1. LES FACTEURS INFLUENÇANT LA REPRÉSENTATION EN MÉMOIRE DES OBJETS VISUELS

Récemment, il a été mis au point des paradigmes permettant d'étudier quels sont les objets de la scène qui sont traités en détail et représentés en mémoire, autant du point de vue de la qualité de leur représentation et de leur rapidité d'accès : le paradigme « flicker » et le paradigme « mudsplash ». Ces deux paradigmes permettent une perception quasi continue de la scène. Le paradigme « flicker » (Rensink, 2002; Rensink et al., 1997, 2000) consiste à présenter, rapidement et en alternance, les versions pré- et post-changement d'une scène visuelle, entre lesquelles est intercalé un masque visuel d'une durée de 80 ms. Le cycle scène pré-changement – masque – scène post-changement est répété jusqu'à ce que les sujets signalent un changement ou que 60 secondes se soient écoulées, comme le montre la Figure 6. Ainsi, le paradigme donne l'impression d'une image clignotante, perçue de manière continue.

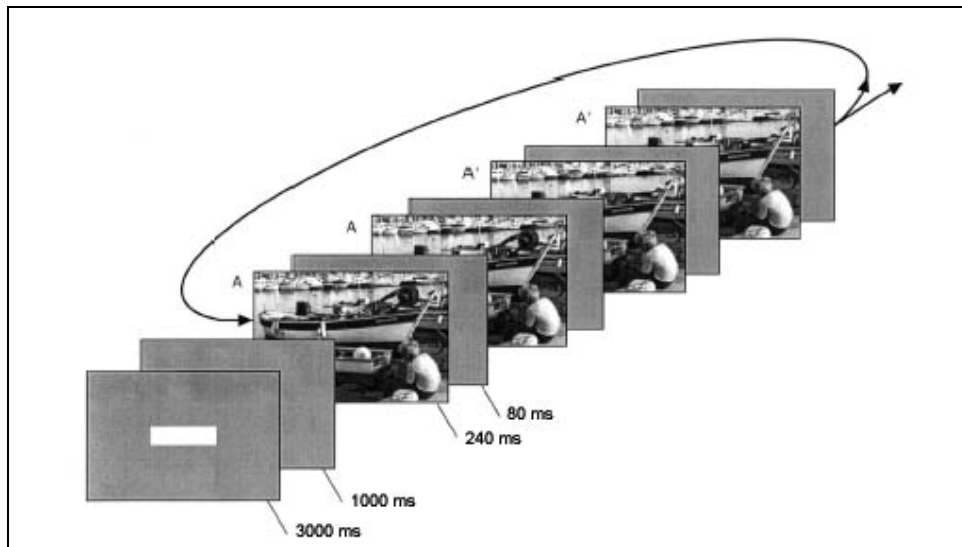


Figure 6 : Paradigme « Flicker ». Les versions originale et altérée de la scène sont alternativement présentées et entrecoupées d'un masque : le cycle s'arrête quand le participant répond ou après qu'une minute s'est écoulée (extrait de Rensink et al., 2000).

Le paradigme « mudsplash » consiste à procéder au changement sur la scène sans qu'il y ait d'interruption visuelle de la scène : dans ce paradigme, l'attraction exogène de l'attention par le changement est court-circuitée par l'apparition concomitante de plusieurs tâches qui recouvrent très brièvement certaines parties de l'image, sans jamais masquer le changement, à la façon des tâches de boue qui parsèment le pare-brise d'une automobile par temps de pluie (d'où le nom du paradigme).

L'avantage de ces deux paradigmes est de permettre d'appréhender la détection des changements, et donc la représentation en mémoire, des objets visuels de la scène « en temps réel », sans que l'on puisse invoquer d'effet d'effacement en mémoire dus aux mouvements oculaires ou à la survenue d'un masque visuel (Wolfe, 1999). Les études, ayant utilisé l'un ou l'autre de ces paradigmes pour appréhender la représentation des scènes visuelles naturelles, ont permis de mettre en évidence un facteur essentiel de la représentation des scènes visuelles en mémoire en condition de vision passive, c'est-à-dire lorsque les participants reçoivent comme instruction de regarder la scène, et éventuellement de chercher les changements qui y surviennent : l'intérêt sémantique de l'information.

1.1. Intérêt sémantique de l'information

Les recherches ayant étudié la représentation de scènes visuelles sous l'angle de la vision passive, et dont une majorité utilise les paradigmes « flicker » ou « mudsplash », ont révélé que les changements étaient différenciellement détectés selon qu'ils survenaient sur des objets d'intérêt central (*ie.*, sur des objets de haut intérêt sémantique qui permettent d'appréhender la scène et d'en extraire la signification) ou sur des objets d'intérêt marginal (*ie.*, sur des objets de faible intérêt sémantique qui ne permettent que peu, voire pas du tout, d'appréhender la scène et d'en extraire la signification). En effet, les changements réalisés sur

des objets d'intérêt central sont massivement mieux détectés, en termes de vitesse comme de quantité, que les changements réalisés sur des objets d'intérêt marginal (Mitroff & Simons, 2002; O'Regan, 2001; O'Regan et al., 2000; O'Regan et al., 1999; Rensink et al., 1997, 2000; Simons & Levin, 1997). Ainsi, il semblerait que les objets d'intérêt central, qui sont aussi les objets qui reçoivent le plus de fixations oculaires lorsque l'exploration visuelle de la scène par les participants est laissée libre (voir chapitre 2), sont les plus susceptibles d'être focalisés par l'attention et représentés en mémoire, de façon suffisamment robuste pour permettre la détection de leur changement. Au contraire, il semblerait que les changements réalisés sur les objets d'intérêt marginal, qui sont les objets les moins enclins à être fovéalisés lorsque l'exploration visuelle de la scène est laissée libre (voir chapitre 2), ne sont que très peu détectés (et, lorsqu'ils le sont, de longues latences de réponse), ce qui suggère que ces objets ne sont que très minoritairement focalisés par l'attention et peu représentés en mémoire. Bien évidemment, il n'est pas rare que les objets d'intérêt central soient, d'une manière générale, plus saillants du point de vue perceptif que les objets d'intérêt marginal (ils sont plus gros, plus lumineux...), ce qui pourrait modérer le rôle de la saillance cognitive dans la représentation de l'information visuelle en mémoire (Scholl, 2000). Dans une récente étude, Itti et Arbib (2005) demandent à des participants de rapporter oralement les éléments d'intérêt central²⁸ de plusieurs scènes dynamiques (les patterns des fixations oculaires sont également recueillis). Simultanément, ils utilisent un modèle computationnel bottom-up de l'orientation de l'attention afin de déterminer les cartes de saillances pour chacune des scènes visuelles utilisées. La comparaison des verbalisations et de la distribution des fixations oculaires des participants avec les zones désignées les plus saillantes par le modèle bottom-up révèle que les éléments centraux du point de vue des verbalisations et des mouvements oculaires

²⁸ Dans l'expérience, les termes « objets les plus importants » sont utilisés

humains sont aussi les plus perceptivement saillants du point de vue bottom-up. Ainsi, il est tout à fait possible que l'avantage représentationnel donné aux objets d'intérêt central ne résulte pas d'un effet de la saillance cognitive (les objets de haut intérêt sémantique sont préférentiellement focalisés par l'attention et représentés en mémoire, alors que les autres objets sont ignorés des traitements attentionnels), mais tout simplement d'un effet de la saillance visuelle (les objets les plus perceptivement saillants sont préférentiellement focalisés par l'attention et représentés en mémoire, alors que les autres objets sont ignorés des traitements attentionnels).

Afin de départager les effets imputables à la saillance visuelle et à la saillance cognitive des objets dits d'intérêt central, Kelley et collaborateurs (Kelley, Chun, & Chua, 2003) ont systématiquement introduit deux changements dans des scènes visuelles complexes, l'un altérant un objet d'intérêt central, l'autre un objet d'intérêt marginal, tout en veillant à ce que les deux changements soient de même saillance visuelle des points de vue de la couleur, de la taille, du contraste avec l'arrière plan et de l'excentricité par rapport au centre de l'image. La tâche des sujets est de détecter le plus rapidement possible un changement entre les deux images (un paradigme « flicker » est utilisé). Selon les auteurs, si la faveur représentationnelle donnée aux objets d'intérêt central ne s'explique que par la supériorité de leur saillance perceptive sur celle des objets d'intérêt marginal, alors ils devraient observer que les participants rapportent autant de changements centraux²⁹ que marginaux³⁰. Le cas échéant, la détection majoritaire des changements centraux renforcerait la thèse de l'intérêt sémantique en tant que facteur de la représentation des informations visuelles en mémoire. Les résultats montrent qu'en dépit de saillances visuelles comparables, les participants rapportent davantage de changements centraux que de changements marginaux, ce qui suggère que la

²⁹ Un changement central altère un objet d'intérêt central

³⁰ Un changement marginal altère un objet d'intérêt marginal

saillance cognitive facilite la représentation en mémoire des objets de la scène. De plus, lorsqu'ils renversent les images présentées aux participants, ce qui a pour conséquence de réduire considérablement la signification des scènes et de fait la saillance cognitive des objets d'intérêt central, Kelley et collaborateurs observent que la préférence de détection pour les changements centraux passe de 81% en condition normale à 69% quand la scène est présentée sens dessus dessous (pour des résultats similaires quand les versions originale et altérée de la scène sont présentées côte à côte, voir Shore & Klein, 2000). Ainsi, lorsque leur intérêt sémantique est réduit, les objets d'intérêt central sont moins susceptibles d'être représentés en mémoire. Ce résultat assied donc le rôle de la saillance cognitive dans la focalisation attentionnelle et la représentation en mémoire des objets visuels. Le système cognitif semble attribuer une priorité représentationnelle aux objets de la scène qui en permettent l'appréhension : l'intérêt sémantique de l'information favorise sa représentation et son accès conscient en mémoire.

1.2. Pertinence de l'information visuelle par rapport à la tâche

Si la priorité attentionnelle et représentationnelle est donnée aux objets de la scène qui présentent un intérêt sémantique pour les participants, l'on devrait, à plus juste titre, observer que le système cognitif traite de manière préférentielle l'information visuelle pertinente par rapport à la tâche qu'il doit accomplir. En effet, en condition de vision active (*ie.*, lorsque les participants doivent réaliser une tâche sur les scènes), la notion d'intérêt sémantique n'est pas suffisamment forte pour rendre compte de la distribution des traitements attentionnels dans les scènes, et les études de Yarbus (1967) montrent bien que le faisceau attentionnel alloué aux régions d'intérêt sémantique de la scène est détourné vers d'autres régions selon la question à laquelle l'exploration visuelle de la scène doit permettre de répondre. Il est donc légitime de

supposer que, de manière comparable, une même scène sera représentée différemment en mémoire selon la tâche qui en guidera l'exploration.

Dans une expérience, Wallis et Bülthoff (2000) placent des participants en simulateur de conduite, soit en tant que conducteur soit en tant que passager du véhicule, et enregistrent leur aptitude à détecter le changement de blocs disposés çà et là dans la scène visuelle, soit sur la route, soit près de la route, soit loin de la route (les réponses correctes sont recueillies). Les résultats montrent une interaction significative du facteur localisation du bloc changé (sur, près, loin de la route) avec le facteur position du sujet (conducteur, passager) sur les performances : alors que l'on n'observe pas de différence de détection quelle que soit la position du bloc changé quand les participants ont la qualité de passager (condition vision passive de la scène), lorsqu'ils ont la qualité de conducteur (condition vision active de la scène), ils perçoivent significativement mieux le changement des blocs situés sur la route que le changement des autres blocs. Ces résultats indiquent, en premier lieu, que la tâche de conduite a entraîné le rétrécissement du champ visuel des participants « conducteurs » à la route seulement, si bien que ne sont traités que les éléments pertinents pour la conduite. En second lieu, ils indiquent que ce rétrécissement ne peut être que le fait de la tâche puisque les blocs situés au-delà du complexe routier et ignorés des participants conducteurs ne sont pas différents, du point de vue perceptif, des blocs qu'ils ont focalisés et représentés.

Dans un même registre, Shinoda et collaborateurs (Shinoda et al., 2001) ont étudié la propension des participants, placés dans un simulateur de conduite, à détecter le changement d'un panneau routier « Interdiction de stationner » en panneau « Stop » (le panneau est placé dans la région d'une intersection avec une autre route), ce changement n'étant visible que pendant une courte durée seulement. Les participants reçoivent pour consigne soit de suivre le véhicule qui les précède, soit, en plus de suivre le véhicule qui les précède, de respecter les

règles habituelles de conduite. La détection des changements est mesurée implicitement, via la durée des fixations oculaires sur la région du changement, et explicitement via le report verbal du changement et le comportement éventuel d'arrêt du véhicule, au croisement où l'exige le panneau « stop ». Les résultats indiquent que la détection du changement du panneau routier dépend de la consigne donnée aux participants. Les participants qui ont reçu la consigne de respecter les règles de conduite *et* de suivre le véhicule qui les précède passent plus de temps à déployer leur regard dans la zone de l'intersection, rapportent davantage de changement et sont plus enclins à stopper leur véhicule au croisement que les autres participants, qui doivent seulement calquer le déplacement de leur véhicule sur celui qu'ils suivent. De plus, il est intéressant de remarquer que la saillance bottom-up peut difficilement rendre compte de ces résultats, du moins pour l'orientation répétée des mouvements oculaires en direction du panneau « stop », puisque des stimuli de plus grande saillance perceptive (par exemple les autres véhicules) ne peuvent pas s'offrir le luxe d'une telle récurrence du regard. Ainsi, les objets visuels présents dans une même scène sont perçus différemment selon les demandes de la tâche. Les exigences de la tâche conduisent donc, non seulement à des différences dans l'allocation du regard et de l'attention dans une même scène visuelle (Yarbus, 1967), mais surtout à des représentations visuelles différentes impliquant différentes informations de la scène.

Dans l'optique d'étudier plus en détail l'influence des exigences immédiates de la tâche sur le traitement et la représentation de l'information visuelle, Hayhoe, Bensiger et Ballard (1998) ont étudié comment est détecté un même changement lorsqu'il survient à différents moments dans la résolution d'une tâche. La tâche à accomplir est une copie de blocs colorés, dans laquelle les participants doivent reconstituer, dans une aire d'assemblage, une configuration modèle, en piochant les blocs adéquats dans une réserve (voir Figure 7). A un moment donné de la tâche, un changement survient sur la couleur du bloc à copier. Ce

changement est réalisé soit *avant* que le bloc à copier soit « pioché » dans la réserve (à gauche dans la Figure 7), soit *après* que ce bloc ait été « pioché » dans la réserve (à droite dans la Figure 7). Étant donné que les participants tendent à découper la tâche de copie du bloc (copier la couleur *et* la position du bloc) en sous tâches élémentaires (copier la couleur *puis* la position du bloc), ce qui leur permet de réaliser la tâche en en allégeant le plus possible la charge mentale (Ballard et al., 1995), les auteurs supposent que le changement de couleur ne devrait réellement être détecté que lorsqu'il survient après « piochage », sans qu'il y ait de détection du changement avant « piochage ». En d'autres termes, les auteurs supposent que le changement de la couleur du bloc sera différenciellement détecté selon qu'il survient *avant* le piochage du bloc, quand la couleur de celui-ci n'a pas encore été encodée, ou *après* le piochage du bloc, quand elle l'a été et qu'elle est désormais utile pour savoir où déposer le bloc dans l'aire d'assemblage. Les durées des fixations oculaires sur le bloc en cours de copie sont recueillies : si le changement est détecté, alors il devrait interférer avec la tâche et cette interférence devrait se manifester par de plus longues fixations sur le bloc à copier.

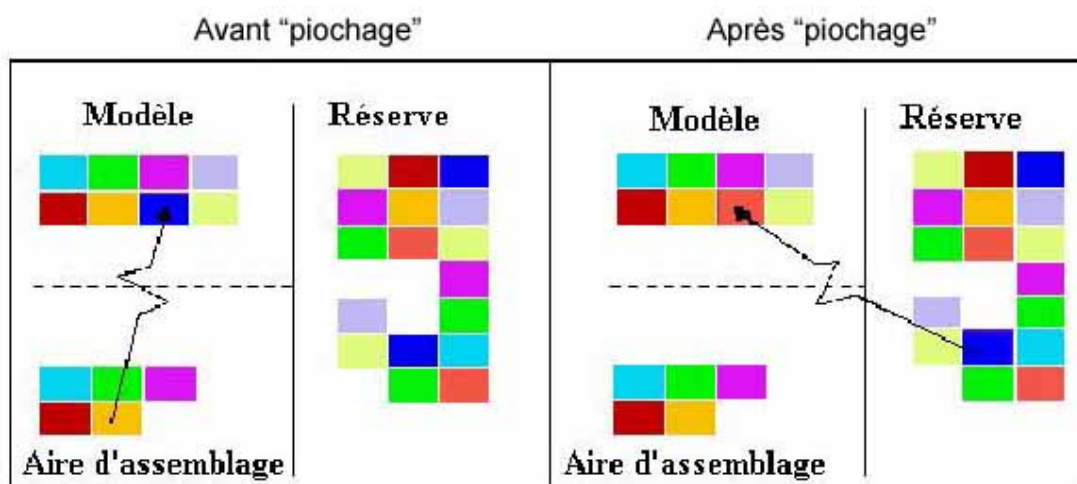


Figure 7: Schéma des deux conditions de l'expérience (Hayhoe et al., 1998). La flèche indique que le changement de la couleur du bloc à copier change pendant une saccade oculaire, à gauche, de l'aire d'assemblage vers le modèle (condition changement du bloc *avant* piochage du bloc copié), et à droite, de la réserve vers le modèle (condition changement du bloc *après* piochage du bloc copié).

Les résultats indiquent que les fixations oculaires sur le bloc du modèle en cours de copie, moyennées sur tous les sujets et tous les essais, sont ralenties de 43 ms par rapport à la condition contrôle (dans laquelle il n'y a pas de changement) dans la condition « changement avant piochage » ; et de 104 ms par rapport à la condition contrôle, dans la condition « changement après piochage »³¹. Ainsi, comme postulé par les auteurs, le bloc à copier n'est représenté en mémoire de façon suffisamment détaillée pour soutenir la détection du changement de sa couleur que lorsque l'information « couleur du bloc » est directement requise par la tâche de copie, c'est-à-dire après piochage du bloc (*ie.*, quand cette information est cruciale pour déterminer où poser le bloc coloré). Ce premier résultat suggère que, même quand la tâche à réaliser et le matériel sont simples, le système cognitif tend à représenter seulement l'information pertinente par rapport à la tâche, au moment où elle est nécessaire. Ainsi, tout comme les traitements attentionnels en condition de vision active sont alloués aux éléments de la scène qui sont pertinents par rapport à la tâche au fur et à mesure qu'ils sont requis (voir chapitre 2), les éléments de la scène semblent être représentés en mémoire uniquement lorsqu'ils sont pertinents par rapport à la tâche et immédiatement nécessaires à sa résolution.

Enfin, Hayhoe et collaborateurs (1998) ont tenté d'examiner dans quelle mesure un changement, particulièrement saillant du point de vue perceptif, mais n'altérant pas les besoins immédiats de la tâche (c'est-à-dire le bloc en cours de copie) était détecté. Pour cela, ils ont procédé, pendant une saccade oculaire vers le modèle (avant ou après piochage), au changement de la couleur de *tous* les blocs du modèle, sauf de celui en cours de copie. Les résultats indiquent que les fixations oculaires qui s'ensuivent sont significativement plus

³¹ Les analyses intra-sujet révèlent que l'on n'observe de différence entre les durées de fixations mesurées en condition expérimentale et contrôle uniquement lorsque le changement survient *après* « piochage » (cette différence est significative pour 4 des 5 sujets de l'expérience) ; aucune différence n'apparaissant quand le changement survient *avant* « piochage ».

longues que dans une condition contrôle dans laquelle il n'y a pas de changement, que ce changement ait pris place avant ou après le « piochage » du bloc en cours de copie. De prime abord, ce résultat peut suggérer que les participants ont quelque part représenté l'ensemble de la scène, indépendamment de la pertinence par rapport à la tâche des éléments qu'elle contient, ce qui leur permet de se rendre compte du changement essuyé par tous les blocs du modèle. Cette hypothèse n'est cependant cohérente ni avec les verbalisations a posteriori des participants, qui rapportent n'avoir remarqué que le changement d'un *seul* bloc sur les *sept* réellement modifiés, ni avec le fait que, dans la condition précédente, le changement d'un seul bloc n'est réellement détecté qu'*après* piochage. En effet, dans le premier cas, l'incapacité des participants à reporter le changement de plus d'un bloc semble peu compatible avec l'idée d'une représentation détaillée de l'ensemble du dispositif visuel, et dans le second cas, la capacité des participants à ne reporter le changement d'un bloc qu'à un certain moment de la résolution de la tâche semble également peu compatible avec l'idée d'une représentation détaillée de l'ensemble du dispositif visuel. Une alternative à cette première explication est de considérer que ce résultat indique, en fait, que les participants ont représenté certaines informations concernant l'arrangement spatial des objets dans la scène, représentation approximative de la scène dont on sait qu'elle sert de cadre de référence au déploiement de l'attention. Cette connaissance sur l'arrangement global de la scène serait suffisante pour permettre aux participants de détecter qu'un changement est survenu dans la condition « changement global » (qui altère l'arrangement du dispositif visuel), et insuffisante pour leur permettre de détecter un changement unique (qui ne modifie pas l'arrangement global du dispositif visuel), dès lors que ce changement unique n'altère pas leurs besoins immédiats. Ainsi, plutôt que de s'expliquer par l'existence d'une représentation détaillée de l'ensemble de la scène, la détection implicite du changement global témoignerait de l'intervention d'épi-représentations précocement activées et stratégiquement utilisées pour guider efficacement le

déploiement de l'attention focalisée vers les éléments de la tâche qu'il s'agit de traiter plus en détail.

Bien entendu, il reste la possibilité que l'extrême saillance perceptive du changement global ait servi de signal d'alerte, et que cette forte stimulation bottom-up ait conduit à la détection implicite du changement. Mais cette possibilité reste, là encore, difficilement interprétable en terme de représentations visuelles détaillées et affranchies du joug de la pertinence par rapport à la tâche. Au contraire, elle renforce l'hypothèse que le facteur « pertinence par rapport à la tâche » est tel que son emprise sur la détection des changements de la scène ne vacille que lorsque les facteurs bottom-up sont extrêmement saillants, puisque la majorité des verbalisations a posteriori ne rend pas justice à l'ampleur du changement réellement effectué.

Cela étant, il reste difficile d'appréhender la question du niveau de détail porté par les représentations de scènes visuelles complexes, comme en témoigne le chapitre suivant.

2. LES REPRÉSENTATIONS VISUELLES : ÉPARSES OU DÉTAILLÉES ?

Le débat concernant le détail de nos représentations visuelles continue de faire rage dans la littérature scientifique, entre d'un côté les partisans de représentations visuelles de la scène limitées au codage détaillé de certains objets visuels seulement, conformément aux limitations de la mémoire visuelle de travail (Ballard et al., 1995; Beck & Levin, 2003; Becker & Pashler, 2002; Dennet, 1993; Hayhoe, 2000; Irwin & Zelinsky, 2002; Noë et al., 2000; O'Regan, 2001; Rensink, 2000a, 2000b; Shinoda et al., 2001; Tatler et al., 2003) ; et de l'autre les partisans d'une représentation complète et détaillée de l'ensemble des objets de la scène,

qui invoquent l'intervention de la mémoire à long terme en plus de la mémoire de travail dans la perception de scènes visuelles naturelles (Henderson & Hollingworth, 2003; Hollingworth, 2004, sous presse-a, sous presse-b, sous presse-c; Hollingworth & Henderson, 2002; Hollingworth et al., 2001; Scott-Brown et al., 2000; Simons et al., 2002; Standing, 1973; Thornton & Fernandez-Duque, 2000).

2.1. Beaucoup d'appelées, peu d'élues : rares sont les informations visuelles à être représentées en mémoire

Une des explications au phénomène de cécité au changement consiste à proposer que, contrairement à notre phénoménologie qui nous assure de représentations visuelles calquées sur le monde réel (Dennet, 1993 ; Noë & O'Regan, 2000; Noë et al., 2000) et laisse croire que nous serions à même de détecter le moindre changement apporté à une scène visuelle (Levin, Drivdahl et al., 2002; Levin et al., 2000), notre incapacité à détecter des changements de grande amplitude signe l'existence de représentations visuelles sommaires et éparées. Ces représentations ne coderaient qu'une petite quantité d'informations visuelles, bien que, à chaque nouvelle fixation, une représentation détaillée de la scène soit créée, mais que, labile et transitoire, elle ne puisse résister aux saccades oculaires. Ainsi, dans la lignée de la théorie de l'intégration de caractéristiques basiques de la scène par le focus de l'attention (Treisman & Gelade, 1980), et de la théorie de l'intégration transaccadique (Irwin, 1992), la représentation en mémoire de l'information visuelle requerrait l'intervention de l'attention focalisée pour résister aux interruptions saccadiques et être maintenue en mémoire sous la forme de « fichiers d'objets » (Kahneman et al., 1992), lesquels constitueraient les seules structures capables de persister pendant quelques secondes (Noles et al., 2005). La représentation d'une scène visuelle naturelle serait donc dépendante des traitements

attentionnels, lesquels seraient guidés par les connaissances sur la scène (résumé, schéma, arrangement spatial et contexte) et les exigences de la tâche. La focalisation de l'attention sur certains objets de la scène permettrait donc leur représentation mémoire, au détriment des autres objets de la scène (Intraub, 1997) : les objets de la scène les plus susceptibles d'attirer le focus de l'attention seraient les plus à même d'être représentés en mémoire (Rensink et al., 1997).

En accord avec ces hypothèses, nombreuses sont les recherches ayant pu mettre en évidence la représentation sélective des objets de la scène en fonction de leur pertinence sémantique : les objets d'intérêt central sont, en règle générale, mieux représentés que les objets d'intérêt marginal, autant du point de vue de la qualité que de la rapidité d'accès à leur représentation (Kelley et al., 2003; Mitroff & Simons, 2002; O'Regan, 2001; O'Regan et al., 2000; O'Regan et al., 1999; Rensink et al., 1997, 2000; Simons & Levin, 1997). De la même manière, les études sur la représentation de scènes visuelles complexes en situation de vision active, c'est-à-dire lorsque l'exploration visuelle de la scène est dévolue à l'exécution d'une tâche particulière (tâche de conduite, Shinoda et al, 2001; Wallis & Bülthoff, 2000 ; de copie de blocs colorés, Hayhoe et al, 1998...), ont, elles aussi, révélé que le système cognitif ne tendait qu'à représenter une quantité limitée d'informations visuelles, les informations pertinentes par rapport à la tâche en cours de réalisation (Ballard et al., 1995; Ballard et al., 1998; Hayhoe, 2000; Hayhoe et al., 2002; Hayhoe et al., 1998; Hayhoe et al., 2003; Karn & Hayhoe, 2000; Pelz et al., 2001; Rao et al., 2002; Shinoda et al., 2001; Triesch et al., 2003; Wallis & Bülthoff, 2000). Ainsi, les buts de la tâche à satisfaire définiraient quelles informations de la scène doivent être préférentiellement traitées, et conduiraient à l'émergence de représentations visuelles de la scène dédiées à la satisfaction de la tâche. Ces représentations seraient prises en charge par la mémoire visuelle de travail, qui permet le

stockage de la localisation et de l'identité des objets pertinents par rapport à la tâche, le temps de la réalisation de la tâche (Melcher, 2001).

Plus généralement, toutes ces recherches postulent que nos représentations visuelles sont éparses et sommaires, et qu'elles ne codent que les informations visuelles focalisées par l'attention et pertinentes pour le système cognitif, soit parce qu'elles facilitent la compréhension de la scène, soit parce qu'elles permettent la résolution d'une tâche. Stockés en mémoire visuelle à court terme, seuls quelques objets visuels (les estimations varient de 1 à 4 - 5 objets, voir Becker & Pashler, 2002 ; Irwin, 1992; Irwin & Zelinsky, 2002; Rensink, 2000a; 2000c), pourraient être représentés de façon suffisamment détaillée pour permettre la détection de leur changement.

A l'extrême, il est postulé que les objets de la scène ne peuvent être représentés qu'isolément, et que par conséquent, nos représentations visuelles ne contiennent que des connaissances schématiques sur la scène, activées dès que celle-ci est présentée (le résumé, l'arrangement spatial et le schéma de la scène), agrémentées du détail d'un objet seulement, celui qui est sous le faisceau de l'attention focalisée (Rensink, 2000a). Ainsi, dans cette perspective, il n'y aurait pas de nécessité à représenter davantage d'information que celle qui est focalisée par l'attention puisque la scène elle-même constituerait une réserve mnésique, dans laquelle le détail de l'information serait directement accessible par un simple mouvement de l'œil ou de l'attention (O'Regan, 1992, 1994, 2001). Bien évidemment, l'hypothèse qui pose le monde en tant que mémoire externe n'est pas limitée à des représentations unitaires des scènes visuelles naturelles. Aussi, il y a fort à parier qu'un système qui représenterait 4 à 5 objets visuels continuerait de tirer parti de l'accessibilité des informations visuelles du monde par des mouvements oculaires et attentionnels fréquents.

2.2. Des représentations visuelles détaillées, qui se rapprochent de la réalité visuelle

Pour certains auteurs, il serait déraisonnable de considérer que nous disposons de représentations visuelles limitées, sur le simple argument que le report verbal du changement ne laisse entrevoir qu'une faible aptitude à détecter les changements réalisés sur des scènes visuelles. En effet, si l'on utilise des mesures implicites de détection des changements (durée des fixations oculaires sur la région changée, tests de choix forcé...), l'on observe que les participants détectent plus de changements qu'ils ne sont capables d'en rapporter oralement (Angelone et al., 2003 ; Fernandez-Duque & Thornton, 2000; Landman et al., 2003; Levin, Simons et al., 2002; Mitroff, Simons, & Levin, sous presse; Simons et al., 2002; Thornton & Fernandez-Duque, 2000). Ces résultats ont conduit à reconsidérer la capacité de codage des représentations visuelles, qui seraient plus détaillées que ne le laisserait supposer la mesure des détections explicites du changement³². Par exemple, il est envisagé que la cécité au changement pourrait se manifester en dépit de représentations visuelles hautement détaillées, et que l'incapacité des sujets à rapporter les changements témoignerait en fait de leur incapacité à procéder à de multiples comparaisons entre les informations de la scène pré-changement représentée en mémoire et les informations véhiculées par la scène post-changement affichée sur l'écran. Dans cette perspective, la cécité au changement ne serait plus que la manifestation de la capacité limitée du système cognitif à procéder à de multiples comparaisons, phénomène connu sous le terme « cécité de comparaison » (Angelone et al., 2003; Landman et al., 2003 ; Levin, Simons et al., 2002; Mitroff et al., sous presse; Scott-Brown et al., 2000; Simons et al., 2002).

³² Il a toutefois été montré (Mitroff et al., 2002) que des explications alternatives à celle de la détection implicite du changement peuvent rendre compte des résultats de certaines de ces études (explications alternatives exprimées en terme de processus d'inférence ou de stratégies explicites de détection)

La théorie la plus aboutie pouvant conduire à l'émergence de représentations visuelles hautement détaillées est très certainement la théorie de la mémoire visuelle et de la représentation des scènes visuelles proposée par Hollingworth et Henderson (Hollingworth, 2003, 2004, sous presse-a, sous presse-b, sous presse-c; Hollingworth & Henderson, 2002; Hollingworth et al., 2001). Cette théorie vise notamment à expliquer les performances de détections considérables observées par les auteurs dans des paradigmes de cécité au changement avec mesure implicite de détection (en majorité, des tests de choix forcés entre deux réponses alternatives ou l'enregistrement des mouvements oculaires). Ces performances de détections (entre 26% et 83% de détection selon le changement et le type de mesure recueilli), sont inhabituellement supérieures à celles généralement observées dans les expériences de cécité au changement, et ce, que les changements introduits dans les scènes soient des additions / disparitions (Henderson et al., 2003 ; Hollingworth & Henderson, 2000), des changements d'orientation (Hollingworth, 2003 ; 2004; Hollingworth & Henderson, 2000), ou des substitutions d'objets de niveau symbolique similaire (« token changes », Hollingworth, 2003 ; 2004; Hollingworth & Henderson, 2002; Hollingworth et al., 2001) ou différent (« type changes », Henderson et al., 2003; Hollingworth & Henderson, 2002).

La théorie de la mémoire visuelle et de la représentation des scènes visuelles proposée par Hollingworth et Henderson postule que la perception des scènes visuelles naturelles est sous-tendue par des représentations visuelles et conceptuelles de haut niveau, relativement détaillées et construites à partir de l'extraction transaccadique d'informations sensorielles. Lorsque l'attention se focalise tour à tour sur divers objets de la scène, sur la base des connaissances précocement activées par la présentation d'une scène visuelle (résumé, schéma, arrangement spatial et contexte) (Henderson, 2003; Henderson & Hollingworth, 1999 ; Henderson et al., 1999), leur détail serait indexé dans des « fichiers d'objets », lesquels

coderaient à la fois la description visuelle (le détail de leurs caractéristiques visuelles) et conceptuelle (leur identité et leur signification) des objets focalisés par l'attention. Dans cette théorie, ces représentations visuelles et conceptuelles des objets les plus récemment focalisés par l'attention seraient activement maintenues en mémoire visuelle à court terme (vMCT), puis consolidées en mémoire à long terme (MLT) (Hollingworth & Henderson, 2002; Hollingworth et al., 2001). Ainsi, avec les multiples fixations oculaires et attentionnelles sur la scène, l'accumulation des informations visuelles et conceptuelles concernant les objets focalisés par l'attention conduirait à la création d'une représentation détaillée de *l'ensemble* de la scène, stockée en mémoire à long terme (Hollingworth, 2004, sous presse-c), représentation dans lesquelles le détail des objets visuels (position, orientation apparence...) serait intimement lié au contexte de la scène (Hollingworth, sous presse-b). Par la suite, quand l'attention se détournerait de l'objet pour en focaliser un autre, les représentations associées déclinerait progressivement en vMCT, mais leur détail resterait néanmoins disponible en MLT. Ainsi, lorsque l'attention se focaliserait à nouveau sur un objet préalablement focalisé, un fort signal de récupération en MLT serait émis et l'information visuelle, focalisée par la nouvelle fixation et actuellement active en vMCT, serait alors comparée à l'information visuelle antérieurement focalisée et récupérée en MLT. Si ces informations sont similaires, elles seraient intégrées et enrichiraient la représentation stockée en MLT; si elles diffèrent, un message d'erreur serait généré et le changement serait détecté.

Aussi, loin d'être éparpées et sommaires, nos représentations seraient visuelles et conceptuelles, et leur détail serait à la fois codé par la mémoire visuelles de travail et la mémoire à long terme.

3. UNE HYPOTHÈSE ALTERNATIVE : PLUSIEURS NIVEAUX DE REPRÉSENTATION DE L'INFORMATION VISUELLE

Bien que visiblement inconciliables, les deux champs théoriques (les représentations des scènes visuelles sont éparées vs. détaillées) partagent de nombreux points communs : la représentation en mémoire des objets de la scène nécessite qu'ils soient (ou aient été préalablement) focalisés par l'attention (Hollingworth & Henderson, 2002; Rensink et al., 1997), la détection consciente des changements est maximale pour les objets les plus récemment fovéalisés (Hollingworth, 2004; Irwin & Zelinsky, 2002), et les connaissances activées précocement (résumé, schéma, arrangement spatial et contexte) interviennent dans la représentation visuelle des scènes naturelles (Henderson & Hollingworth, 1999; Henderson et al., 1999; Intraub, 1997; Rensink, 2000a). En fait, on peut presque considérer que ces champs théoriques ne divergent qu'à propos du niveau de détail porté par les représentations visuelles des scènes naturelles. Or, l'on peut invoquer plusieurs facteurs pour rendre compte de ces estimations divergentes. Dans un premier temps, les durées utilisées dans la présentation des scènes pré-changement sont souvent très différentes et peuvent varier de 240 ms (Rensink et al., 2000) à 20 s (Hollingworth, 2003), et dans certaines études, les participants reçoivent pour consigne de « regarder minutieusement la scène en vue d'un test de mémoire ultérieur » (Hollingworth, 2003; Hollingworth & Henderson, 2002; Hollingworth et al., 2001), deux faits qui peuvent avoir entraîné la mise en place de stratégies de mémorisation des détails de la scène dans un cas, et pas dans l'autre. De plus, les oppositions théoriques sur le détail des représentations visuelles reposent sur l'utilisation de différents niveaux de mesures: lorsque les uns mesurent la part de détection explicite des changements via les reports verbaux des changements perçus consciemment (Kelley et al., 2003; Mondy & Coltheart, 2000; Most et al., 2005; O'Regan, 2001; O'Regan et al., 2000; O'Regan et al., 1999; Rensink, 2000c;

Rensink et al., 1997, 2000; Scholl, 2000; Shinoda et al., 2001; Simons, 1996; Simons & Chabris, 1999; Tatler et al., 2003; Wallis & Bülthoff, 2000; Werner & Bjorn, 2000; Williams & Simons, 2000), les autres envisagent des mesures alternatives, qui rendent compte des détections implicites de changements dont les participants n'ont pas forcément pris conscience (Fernandez-Duque & Thornton, 2000; Hollingworth, 2003, 2004; sous presse-b ; Hollingworth & Henderson, 2002; Hollingworth et al., 2001; Scott-Brown et al., 2000; Thornton & Fernandez-Duque, 2000). Or, il n'est pas impossible que les trois ou quatre objets les plus récemment focalisés par l'attention puissent être les seuls à avoir été suffisamment représentés en mémoire pour que la détection de leur changement soit accessible au report conscient. Dans ce cas, le reste des informations de la scène pourrait être codé à un niveau représentationnel inférieur, inaccessible au report conscient, mais dont le contenu pourrait être mesuré indirectement, grâce à l'utilisation de mesures implicite du changement. Cette hypothèse serait, d'ailleurs, en accord avec la notion de compétition entre les objets visuels pour l'attribution des ressources cognitives limitées.

En effet, on sait que les objets de la scène entrent en compétition pour bénéficier des ressources de traitement et ainsi accéder au niveau de représentation consciente, puisque, à chaque instant, seule une petite quantité du flot d'informations visuelles qui bombarde notre rétine peut être traitée et utilisée pour contrôler nos comportements (Desimone & Duncan, 1995)³³. A l'issue de cette compétition, on trouve (i) les objets qui seront préférentiellement focalisés par l'attention, consolidés en mémoire, et accessibles de manière consciente, (ii) les objets qui bénéficieront de traitements prolongés mais insuffisants à la représentation consciente et (iii) les objets qui seront oubliés, comme le montrent VanRullen et Koch (2003) dans une étude sur la compétition des objets visuels et les représentations différentielles qui

³³ Les déterminants de cette compétition entre les objets visuels pour la représentation consciente se situent probablement entre la saillance bottom-up des objets visuels et leur pertinence par rapport aux comportements à produire (Desimone & Duncan, 1995 ; VanRullen & Koch, 2003; Yantis, 2005)

en découlent. Dans cette étude, ils présentent très brièvement (250 ms) différentes scènes visuelles contenant chacune dix objets visuels (un banc, un vélo, un ballon....) aux participants, chaque scène étant suivie d'un masque visuel. Les auteurs combinent ensuite trois paradigmes expérimentaux différents (rappel libre, reconnaissance via des tests de choix forcé et amorçage négatif)³⁴ sur ces mêmes scènes, dans le but d'observer quels objets de la scène atteignent le niveau de la représentation consciente, lesquels atteignent le niveau de la représentation implicite, et lesquels sont oubliés. Les résultats montrent, dans un premier temps, que les participants ne peuvent rappeler que deux ou trois objets de la scène qui leur a été présentée, ce qui indique que seuls ces quelques objets ont atteint le niveau de la représentation consciente. Dans un second temps, les résultats indiquent que, si l'on a recours à des outils suffisamment sensibles pour mesurer le contenu des représentations implicites (eg., test de choix forcé), les participants sont en mesure de reconnaître deux ou trois objets supplémentaires. Enfin, un fait intéressant émerge lorsque les auteurs testent la possibilité d'un amorçage négatif par les objets de la scène qui n'ont été ni rappelés ni reconnus. Pour ce faire, ils utilisent une tâche d'appariement³⁵, dans laquelle les participants doivent déterminer, le plus rapidement possible, si oui ou non le mot qu'ils ont lu sur l'écran de l'ordinateur correspond à l'image qui y est maintenant affichée. Les auteurs remarquent que le nombre d'erreurs et les latences de réponses aux objets de la scène qui n'ont ni été rappelés ni été reconnus sont supérieurs au nombre d'erreurs et aux latences de réponses observés pour ces mêmes objets dans une situation contrôle, dans laquelle les objets n'ont pas encore été

³⁴ Dans la condition expérimentale, les participants sont confrontés à ces trois paradigmes dans l'ordre suivant : rappel, reconnaissance, amorçage ; alors que dans la condition contrôle, les participants sont confrontés à ces trois paradigmes dans l'ordre inverse

³⁵ « matching »

présentés dans une scène visuelle³⁶. Ce résultat suggère que la trace de l'activité cognitive et neurale, qui n'a pourtant pas permis de conduire à une représentation de ces objets qui serait accessible par les tests de rappel ou de reconnaissance, serait néanmoins en mesure de modifier (en l'occurrence ici, de ralentir) le traitement de ce même stimulus, lorsqu'il est à nouveau présenté aux participants. Cet amorçage négatif indiquerait donc que ces objets ont été filtrés et activement inhibés par le système cognitif, à un stade tardif du traitement visuel.

En résumé, cette étude de VanRullen et Koch (2003) indique que les différents objets de la scène sont codés par des niveaux de représentation différents, qui sont accessibles par des mesures différentes, et ont des effets différents sur les comportements des participants. Notons par ailleurs que, bien que cette étude soit incapable de rendre compte des règles régissant l'entrée des objets visuels dans tel ou tel niveau de représentation, quelques éléments de réponse sont apportés par les expériences sur la cécité au changement, qui suggèrent que les objets de la scène pertinents du point de vue de la compréhension de la scène ou de la résolution d'une tâche s'avèrent les plus à même d'accéder aux niveaux supérieurs de la représentation en mémoire.

³⁶ Rappelons que dans cette condition contrôle, l'ordre de passation des paradigmes est inversé par rapport à la condition expérimentale

– CHAPITRE QUATRE –

PROBLÉMATIQUE

L'objectif de ce travail de thèse est d'évaluer l'influence de la tâche sur la représentation des scènes visuelles en mémoire.

Grâce aux recherches récemment entreprises dans le domaine de la détection de changements et la représentation des scènes visuelles, il est aujourd'hui communément admis que les informations visuelles ont d'autant plus de chances d'être focalisées par l'attention (Hayhoe, 2000; Hayhoe et al., 2002; Land & Hayhoe, 2001; Triesch et al., 2003; Yarbus, 1967) et représentées en mémoire (Ballard et al., 1995; Ballard et al., 1998; Hayhoe et al., 1998; Hayhoe et al., 2003) qu'elles sont pertinentes par rapport à la tâche que nous devons réaliser sur les scènes. La prégnance de ce facteur fonctionnel est telle qu'elle peut outrepasser les effets de la saillance perceptive qui sont pourtant connus pour être robustes (Theeuwes, 1994, 2005; Theeuwes et al., 1998; Theeuwes et al., 2004), et cette prédominance a pu être observée dans des tâches de recherche visuelle, de poursuite visuelle, ou de conduite automobile (Folk et al., 1992 ; Most et al., 2005 ; Shinoda et al., 2001). Or si la pertinence de l'information visuelle est un facteur suffisamment puissant pour s'affranchir de l'influence de la saillance perceptive, il reste difficile de distinguer dans quelle mesure elle parvient également à s'affranchir du joug de la saillance sémantique de l'information. Par saillance sémantique, nous faisons référence à l'intérêt sémantique qu'un objet visuel représente pour la compréhension la scène, intérêt dont les recherches sur la cécité au changement en

condition de vision passive (*ie.*, lorsque les participants reçoivent pour consigne de regarder l'image) ont montré qu'il constituait également un facteur de la représentation en mémoire des scènes visuelles, suffisamment robuste, lui aussi, pour surpasser de simples effets de saillance perceptive (Kelley et al., 2003). En fait, et comme le montrent les chapitres précédents, les objets visuels portés par une scène visuelle sont voués à un futur incertain : selon leur aptitude à capter le faisceau attentionnel (du fait de leur saillance cognitive ou perceptive) et en vertu de la compétition qui prend place entre eux pour la répartition des ressources cognitives qui sont limitées (Desimone & Duncan, 1995; Karn & Hayhoe, 2000 ; Luck & Vogel, 1997; VanRullen & Koch, 2003), les différents objets de la scène vont (i) soit bénéficier de traitements visuels et attentionnels soutenus, qui leur permettront d'accéder aux plus hauts niveaux de représentation en mémoire, d'être disponibles au report conscient et de guider les comportement ultérieurs, (ii) soit bénéficier de traitements moins appuyés, qui s'ensuivront d'un codage au sein de représentations de niveau inférieur en mémoire non accessibles via des moyens conscients, mais qui leur permettra néanmoins d'influencer nos comportements, (iii) soit être activement inhibés et oubliés. Toutefois, s'il reste difficile de prédire leur avenir représentationnel, des arguments nous permettent d'avancer que les objets de la scène cognitivement saillants, du fait de leur intérêt sémantique ou de leur pertinence par rapport à la tâche (Hayhoe et al., 1998; O'Regan et al., 2000; Pani, 2000; Rensink, 2000b, 2002; Rensink et al., 1997; Shinoda et al., 2001), occuperont les plus hautes sphères représentationnelles (Maunsell, 1995; VanRullen & Thorpe, 2001b).

L'hypothèse selon laquelle il existe différents niveaux de représentation de l'information visuelle n'est pas nouvelle dans la littérature. En effet, nombreux sont les modèles qui postulent deux niveaux de traitements et de représentation des informations portées par une scène visuelle, un niveau préattentif qui administre la question de la saillance

perceptive, et un niveau attentif qui gère les traitements de plus haut niveau et permet, entre autres, la reconnaissance des objets (Duncan, 1984; Irwin, 1992; Kahneman et al., 1992; Treisman & Gelade, 1980; Ward & Duncan, 1996; Wolfe, 1994). Par ailleurs, dans son excellente revue de la question, Maunsell (1995) propose que les premières étapes du traitement visuel fournissent une représentation fidèle et labile du monde réel (voir aussi Rensink, 2000a) alors que les étapes ultérieures sont dévolues au traitement approfondi des informations visuelles d'intérêt pour celui qui regarde la scène. Ainsi, en filtrant activement les informations non pertinentes et en accentuant celles qui doivent être conservées en mémoire, le cerveau crée une représentation dynamique du monde visuel, qui change avec les besoins du sujet percevant. Aussi alléchante que soit cette conception, elle laisse en suspens une question cruciale : quelles informations visuelles considérer comme « pertinentes » et donc susceptibles d'accéder aux niveaux supérieurs de la représentation ? Doit-on considérer comme « pertinentes » les informations de haut intérêt sémantique, qui facilitent la compréhension d'une scène visuelle, dont on sait qu'elles sont représentées en mémoire dans des conditions de vision passive, ou bien doit-on considérer comme « pertinentes » les informations qui sont pertinentes par rapport à la tâche, dont on sait qu'elles sont particulièrement représentées en mémoire en condition de vision active ? Jusqu'à présent, aucune étude connue n'a réellement examiné comment ces deux types d'informations (intérêt sémantique et pertinence par rapport à la tâche) sont représentés en mémoire, et les modèles de l'attention visuelle, qui pourraient apporter des éléments de réponse, conduisent à des prédictions contrastées.

Selon le modèle de l'architecture triadique de Rensink (2000), qui stipule que les représentations visuelles ont une capacité limitée à un seul objet (celui qui se trouve actuellement sous le faisceau attentionnel), on peut prédire qu'un objet aura autant de chances d'être représenté en mémoire qu'il aura de chances d'être focalisé par l'attention. Aussi, tout

mène à penser que l'objet qui sera immédiatement pertinent pour la tâche en cours de résolution et donc le plus susceptible de s'octroyer l'exclusivité des traitements attentionnels et visuels (Ballard et al., 1995; Shinoda et al., 2001; Triesch et al., 2003; Yarbus, 1967) sera l'objet représenté en mémoire.

Les postulats des modèles computationnels (Navalpakkam et al., 2005; Navalpakkam & Itti, 2002, 2005; Oliva et al., 2003; Torralba, 2003) nous conduisent à des prédictions plus nuancées. Dans l'ensemble, ces modèles proposent que l'attention sera allouée aux zones d'intérêt de la carte de saillance, laquelle carte intègre les demandes de la tâche, la saillance sémantique (déterminée par le contexte, le schéma de scène, le résumé ou l'arrangement spatial), et la saillance perceptive des différents objets de la scène. Ainsi, ces modèles postulent que le focus de l'attention sera guidé en priorité vers les objets pertinents par rapport à la tâche, de sorte que ces objets seront rapidement focalisés par l'attention et représentés en mémoire. Cependant, on ne peut pas exclure, à partir de ces modèles, que les objets d'intérêt central seront exempts de traitements attentionnels soutenus, d'autant qu'ils s'avèrent être souvent les objets les plus perceptivement saillants (Itti & Arbib, 2005). Or, il n'est pas possible de considérer que les objets pertinents par rapport à la tâche *et* les objets d'intérêt central seront tous représentés en mémoire, compte tenu de la capacité limitée de la mémoire visuelle de travail (Luck et Vogel, 1997), pas plus qu'il est possible d'ignorer que ces deux types d'objets visuels risquent très probablement d'entrer en compétition pour accéder aux plus hauts niveaux représentationnels.

Cette thèse examine la possibilité que les objets de la scène, cognitivement saillants du fait de leur intérêt sémantique ou de leur pertinence par rapport à la tâche, vont être appréhendés avec des priorités représentationnelles différentes. En effet, cette thèse défend l'idée que notre perception des scènes visuelles est sous-tendue par deux hauts niveaux de

représentation, l'un codant les informations d'intérêt sémantique (niveau de représentation sémantique), l'autre codant les informations pertinentes par rapport à la tâche (niveau de représentation fonctionnel). Plus particulièrement, notre thèse suppose que la pertinence par rapport à la tâche est le facteur qui prédomine la compétition entre les objets de la scène pour la représentation en mémoire. Par conséquent, le point de vue argumenté dans cette thèse est que le niveau de représentation fonctionnel des scènes visuelles prend l'ascendant sur le niveau de représentation sémantique, et qu'il constitue ainsi le niveau le plus accessible consciemment et le plus à même d'influencer nos comportements.

Aussi, cette thèse examine, en second lieu, plus en détail les modalités de la représentation en mémoire des informations pertinentes par rapport à la tâche. En effet, les différents travaux menés dernièrement en condition de vision active (Hayhoe, 2000; Hayhoe et al., 1998; Hayhoe et al., 2003; Shinoda et al., 2001; Wallis & Bühlhoff, 2000) suggèrent que les éléments de la scène pertinents par rapport à la tâche sont extraits, traités et représentés en mémoire au fur et à mesure qu'ils sont nécessaires à la résolution de la tâche. Ceci dit, il reste difficile d'estimer, à partir de ces travaux qui utilisent principalement les mouvements oculaires comme mesure de la représentation en mémoire des informations visuelles, à quel niveau de représentation ces informations sont codées et d'identifier réellement les déterminants de la représentation en mémoire des informations pertinentes par rapport à la tâche. L'idée défendue dans cette thèse est que les informations pertinentes par rapport à la tâche sont codées au sein du niveau fonctionnel de représentation des scènes visuelles, qui est le niveau le plus accessible consciemment. Nous proposons par ailleurs que ce niveau de représentation fonctionnel est de nature dynamique. En effet, notre hypothèse est que les informations codées par ce niveau sont sans cesse actualisées au fur et à mesure que la résolution de la tâche progresse, de sorte que nos représentations visuelles fonctionnelles

constituent une description circonstanciée des informations visuelles immédiatement et directement requises par la tâche.

Les hypothèses défendues dans cette thèse sont éprouvées dans plusieurs études comportementales. Ces études utilisent un paradigme de détection des changements (Pashler, 1988) sur des scènes visuelles complexes afin de permettre d'évaluer quelles informations visuelles ont été focalisées par l'attention et représentées en mémoire (Shapiro, 2000). Dans une première série d'expériences, nous manipulerons deux facteurs, l'intérêt sémantique et la pertinence par rapport à la tâche des objets de la scène, afin montrer que les objets de la scène sont différenciellement représentés en mémoire selon qu'ils sont pertinents par rapport à la tâche ou d'intérêt central pour la compréhension de la scène. L'hypothèse testée dans ces expériences sera que le plus haut niveau de représentation de la scène est fonctionnel (*ie.*, codera les objets pertinents par rapport à la tâche), et que les objets liés à la compréhension de la scène sont codés à un niveau de représentation inférieur. Nous concentrerons ensuite nos recherches sur le niveau de représentation fonctionnel des scènes visuelles. Notre hypothèse sera que les représentations visuelles de niveau fonctionnel sont transitoires et dynamiques puisqu'elles codent de manière préférentielle les objets de la scène immédiatement requis par les exigences de la tâche. Afin de tester cette hypothèse, une seconde série d'expériences utilisera des scènes visuelles complexes présentées en situation dynamique. Cette présentation des scènes visuelles en situation dynamique permettra d'étudier comment la représentation d'une même scène diffère selon l'immédiateté des demandes de la tâche. Notre objectif est de montrer que le contenu du niveau fonctionnel des représentations visuelles évolue avec la résolution de la tâche et qu'il répond à un critère d'immédiateté des besoins fonctionnels. Enfin, une des originalités de ce travail sera de proposer un modèle multinomial de la représentation des scènes visuelles complexes qui permettra de tester plus en avant le

caractère fonctionnel des représentations visuelles de haut niveau, en défendant l'idée que le système cognitif ne code que les objets de la scène juste nécessaires et suffisants pour accomplir la tâche qui lui incombe.

DEUXIÈME PARTIE :

ÉTUDES COMORTEMENTALES

– CHAPITRE CINQ –

EFFET DES FACTEURS SÉMANTIQUES ET FONCTIONNELS SUR LA PERCEPTION DE SCÈNES VISUELLES COMPLEXES

Les expériences exposées dans ce chapitre visent à étudier la représentation de scènes naturelles, et plus particulièrement les facteurs qui influencent la représentation en mémoire des informations portées par une scène visuelle complexe. Les travaux sur la cécité au changement³⁷ menés ces dernières années ont mis en évidence deux facteurs influençant la représentation des scènes visuelles en mémoire : l'intérêt sémantique de l'information visuelle, central ou marginal, selon que l'information visuelle permet de comprendre et de donner du sens à la scène ou non (Mitroff & Simons, 2002; O'Regan, 2001; O'Regan et al., 2000; O'Regan et al., 1999; Rensink et al., 1997, 2000; Simons & Levin, 1997), et la pertinence de l'information visuelle par rapport à la tâche, pertinente ou non selon que l'information est impliquée ou non dans la résolution de la tâche (Ballard et al., 1995; Hayhoe, 2000; Shinoda et al., 2001; Wallis & Bülthoff, 2000). Or, aucune recherche n'a réellement étudié l'influence conjointe de ces deux facteurs sur la perception des scènes naturelles complexes en condition de vision active, c'est-à-dire quand les participants doivent réaliser une tâche sur les scènes visuelles. Les deux expériences présentées dans ce chapitre visent à étudier les effets (respectifs et conjoints) des facteurs « intérêt sémantique de

³⁷ La cécité au changement est, rappelons-le, un paradigme expérimental qui permet de déterminer quelles sont les informations d'une scène visuelle qui ont été focalisées par l'attention et consécutivement représentées en mémoire, en estimant, via la mesure de la détection explicite des changements, l'accès conscient au détail de ces représentations visuelles de la scène.

l'information visuelle » et « pertinence par rapport à la tâche de l'information visuelle » sur la représentation de scènes visuelles en mémoire. Sur ce point, différents modèles se distinguent et conduisent à des prédictions contrastées sur le contenu des représentations visuelles en conditions de vision active. D'un côté, il est postulé que, dès les premières étapes du traitement visuel, le déploiement attentionnel est biaisé par les buts de la tâche, qui guident le focus de l'attention vers les éléments de la scène qui sont directement nécessaires à la résolution de la tâche (Navalpakkam & Itti, 2002, 2005). Étant donné que la représentation en mémoire requiert que l'information ait été focalisée par l'attention (Rensink et al., 1997; Simons & Levin, 1997), les postulats de cette première catégorie de modèles nous conduisent à supposer que les informations visuelles pertinentes par rapport à la tâche étant les plus à même d'être focalisées par l'attention, elles sont également les plus à même d'être représentées en mémoire. D'un autre côté, il est postulé que le traitement attentionnel des informations visuelles pertinentes par rapport à la tâche s'accompagne, quoi qu'il en soit, du traitement attentionnel des informations visuelles impliquées dans la compréhension et l'appréhension de la scène visuelle (Henderson & Hollingworth, 2003; Henderson et al., 1999; Oliva et al., 2003), ce qui nous amène à penser qu'il y a de grandes chances que ces informations soient quelque part représentées en mémoire.

Les deux expériences présentées dans ce chapitre partagent un même paradigme expérimental de cécité au changement. Ce paradigme consiste à présenter, dans un premier temps, une scène pré-changement que les participants doivent explorer soit librement, c'est-à-dire sans aucune contrainte expérimentale autre que « *regarder la scène* » (condition « vision passive »), soit de manière à satisfaire une tâche de recherche visuelle de type « *clicquer sur les voitures* » (condition « vision active »). L'exploration des scènes visuelles est ensuite interrompue par l'apparition d'un masque expérimental d'une durée de 500 ms, suivie par

l'apparition de la scène post-changement : la tâche des participants (quelle que soit la condition expérimentale) est de signaler oralement les changements perçus (en précisant la nature du changement et sa localisation dans la scène). Les changements réalisés sur les scènes résultent du croisement des facteurs intérêt sémantique de l'information visuelle (central *vs.* marginal) et pertinence de l'information visuelle par rapport à la tâche (pertinent *vs.* non pertinent). Dans les deux expériences, la détection des changements est utilisée comme une mesure indiquant quels objets de la scène ont été focalisés par l'attention et représentés en mémoire.

1. EXPÉRIENCE 1 : EFFET DES FACTEURS FONCTIONNELS SUR LA PERCEPTION DE SCÈNES VISUELLES COMPLEXES.

L'expérience 1 vise à appréhender dans quelle mesure l'intérêt sémantique de l'information visuelle et sa pertinence par rapport à la tâche influencent respectivement la détection des changements sur des scènes naturelles complexes, et dans quelle mesure ces deux facteurs conduisent à des post-effets attentionnels. En effet, l'utilisation d'un paradigme de détection de changements en un seul cycle (présentation scène pré-changement - masque - présentation scène post-changement, Pashler, 1988) nécessite, pour qu'il y ait détection des changements, que les objets de la scène pré-changement focalisés par l'attention soient représentés *et* que leur représentation subsiste assez longtemps en mémoire, afin de soutenir la comparaison entre les aspects pré- et post-changement de l'objet changé. Les paradigmes expérimentaux plus traditionnellement utilisés dans les recherches sur la cécité aux

changements (paradigmes « flickers³⁸» ou « mudsplash³⁹») étudient, par la présentation quasi continue de la scène visuelle, les effets en temps réel de la focalisation attentionnelle sur la représentation en mémoire de l'information visuelle.

1.1. Méthode

1.1.1. Participants

Quarante étudiants inscrits en licence de Psychologie à l'Université de Provence ont participé à cette recherche. Tous ont attesté d'une acuité visuelle normale ou corrigée. Vingt-sept de ces participants ont été aléatoirement répartis en deux groupes selon la condition de passation de l'expérience : groupe « vision active » ($n = 16$) et groupe « vision passive » ($n = 11$). Les autres participants ont contribué à contrôler l'intérêt sémantique du matériel ($n = 6$) et la visibilité du changement ($n = 7$). Tous les participants étaient naïfs du point de vue des hypothèses de l'expérience.

1.1.2. Matériel

Vingt-six photographies couleurs (dont cinq utilisées dans la phase d'entraînement) représentant des scènes naturelles d'intérieur et d'extérieur, provenant d'une source

³⁸ Le paradigme « flicker » (Rensink et al., 1997) consiste à répéter très rapidement le cycle image pré-changement (durée 240ms) –masque (durée 80ms) -image post-changement (durée 240ms), de sorte à obtenir une image qui donne l'impression de clignoter très rapidement (d'où le nom du paradigme, « flicker » signifie « clignoter » en anglais).

³⁹ Le paradigme « mudsplash » (O'Regan et al., 1999) consiste à parsemer la scène de plusieurs taches, de la même façon que rouler dans une flaque de boue parsèmerait le pare-brise d'un véhicule de gouttes d'eau boueuse (d'où le nom du paradigme : « mudsplash » signifie « tache de boue » en anglais). Dans ce paradigme, la scène est perçue de façon continue, et les taches ne masquent jamais le changement.

personnelle, ont été numérisées au format 342 x 512 ou 512 x 342 pixels. Toutes les photographies ont été dupliquées et quatorze de ces duplicata ont été numériquement retouchés de façon à obtenir vingt et une paires de photographies, dont 7 paires de photographies strictement identiques et quatorze paires de photographies qui ne diffèrent que par la modification d'un objet de la scène. Les changements réalisés sur les scènes visuelles consistent en l'addition ou la disparition de tout ou partie d'un objet de la scène, comme il est d'usage de procéder dans les paradigmes de détection de changement (Rensink et al., 1997, 2000). Les changements ont été déterminés par la manipulation orthogonale de deux facteurs expérimentaux : la pertinence du changement par rapport à la tâche (pertinent vs. non pertinent) et l'intérêt sémantique du changement (central vs. marginal). Le facteur « pertinence du changement par rapport à la tâche » renvoie à l'implication de l'objet changé dans une tâche de recherche visuelle : un changement pertinent par rapport à la tâche altère un objet qui est la cible de la tâche de recherche visuelle, un changement non pertinent par rapport à la tâche concerne tout autre objet de la scène qui n'est pas la cible de la tâche de recherche visuelle. Le facteur « intérêt sémantique du changement » renvoie, quant à lui, au degré d'implication de l'objet changé dans la compréhension et l'interprétation de la scène visuelle. Un changement qui altère un objet fortement impliqué dans la compréhension et l'interprétation de la scène est dit « d'intérêt central », alors qu'un changement qui altère un objet faiblement impliqué est dit « d'intérêt marginal ». L'intérêt sémantique des objets de la scène a été déterminé préalablement par un test de contrôle, selon le procédé mis en place par Rensink et ses collaborateurs (Rensink et al., 1997). Pour ce faire, nous avons demandé à des observateurs différents de ceux de l'expérience de décrire chacune des photographies présentées en commençant leur description par les objets de la scène qui leur semblaient les plus importants. Les trois éléments les plus fréquemment cités en début de description constituaient les objets d'intérêt central ; ceux qui n'avaient jamais été mentionnés au cours

de la description des photographies constituaient les objets d'intérêt marginal. Il résulte de la manipulation orthogonale des facteurs « pertinence du changement par rapport à la tâche » et « intérêt sémantique du changement » quatre combinaisons de changements : Pertinent Central (PC), Pertinent Marginal (PM), Non Pertinent Central (NPC), et Non Pertinent Marginal (NPM). Le tableau de l'annexe 1 montre la répartition des changements en fonction de leur pertinence et de leur intérêt sémantique.

Afin de nous assurer que les changements introduits dans les scènes étaient d'envergure considérable, et, par conséquent, aisément détectables selon les règles définies par Simon et Levin (1998)⁴⁰, nous avons procédé à un test de contrôle de la visibilité des changements. Dans ce test, nous avons présenté simultanément les photographies pré- et post-changement à des observateurs différents des participants de l'expérience et du test de contrôle de l'intérêt de objets visuels, pendant 10 s : tous les changements ont été détectés et identifiés sans peine par ces observateurs.

Le masque expérimental, du même format que les photographies, est constitué de différentes figures géométriques grisées. Les paires de photographies sont regroupées en trois blocs de 7 paires de photographies selon un ordre aléatoire contrôlé, contraignant chaque bloc à contenir au moins un exemplaire de chaque combinaison de changement (PC, PM, NPC, et NPM) et une paire de photographies identiques. L'ordre de passation des blocs ainsi que l'ordonnancement des paires de photographies à l'intérieur des blocs restent constants pour tous les participants.

⁴⁰ Simons et Levin (1998) stipulent que pour être aisément détectable, un changement doit être facilement verbalisable et discriminable quand les photographies pré- et post-changement sont présentées côte à côte.

1.1.3. Équipement

L'expérience, pilotée par ordinateur, se déroule dans un box expérimental de l'Université de Provence, dans des conditions d'éclairage confortables. Les participants sont installés à une distance approximative de 50 cm de l'ordinateur. L'ordinateur utilisé pour la programmation et passation de l'expérience, ainsi que pour l'enregistrement des données est un macintosh iBook, modèle 116498, équipé d'une souris. Les photographies ont été modifiées avec le logiciel Adobe Photoshop Elements®, et l'expérience a été programmée avec le logiciel Hypercard®.

1.1.4. Procédure : tâche de détection de changements

Avant de commencer l'expérience, une consigne informe les participants que des changements peuvent avoir lieu sur les photographies qui leur seront présentées lors de la phase d'entraînement et lors de l'expérience, et qu'ils devront signaler tout changement dès détection. Les participants ne reçoivent pas d'information concernant le type ou la fréquence d'occurrence des changements opérés sur les scènes visuelles.

L'expérience se déroule en trois phases successives comme le montre la Figure 8. Nous commencerons par décrire le déroulement de l'expérience pour le groupe «vision active» puis nous expliciterons ce en quoi il diffère pour le groupe «vision passive».



Figure 8. Paradigme de cécité au changement utilisé dans les expériences 1 et 2. Un essai commence avec la présentation d'un écran de transition, suivi de la présentation de la scène pré-changement que les participants doivent explorer soit librement (groupe «vision passive») soit de manière à satisfaire une tâche de recherche visuelle de type « cliquer sur ... » (groupe «vision active»). La durée allouée à l'exploration des scènes est soit libre (expérience 1), soit limitée (expérience 2). La scène pré-changement est ensuite remplacée par un masque visuel d'une durée de 500 ms, puis par la présentation de la scène post-changement. La tâche des participants est de signaler tout changement perçu et d'en préciser oralement la nature et la position, soit en temps libre (expérience 1) soit en temps limité (expérience 2). Enfin, les participants sont soumis à un test de détection implicite des changements, dans lequel ils doivent désigner l'objet changé parmi quatre propositions (expérience 2 seulement). Ici, un changement NPC est illustré par l'addition d'une fontaine dans la scène

Tout nouvel essai débute avec la présentation d'un écran de transition. Cet écran, qui vise à préparer le participant à la passation d'un nouvel essai, reste affiché jusqu'à ce que le participant clique sur un bouton « *Prêt* » qui déclenche un nouvel essai. Chaque essai se déroule comme suit : dans un premier temps, la scène pré-changement apparaît sur l'écran : une consigne spécifique, accolée à la scène pré-changement, enjoint aux participants du groupe «vision active» de sélectionner, via un click de souris, l'ensemble des objets de la

scène qui appartiennent à une catégorie spécifiée⁴¹ (par exemple : « *Cliquer sur les fenêtres* »). En d'autres termes, l'exploration des scènes pré-changement par les participants du groupe «vision active» est dévolue à la réalisation d'une tâche de recherche visuelle. Les objets visuels concernés par la tâche de recherche visuelle (dans notre exemple : les fenêtres) constituent les objets pertinents, les autres objets de la scène constituent les objets non pertinents. Chaque click sur un objet pertinent s'accompagne d'un feed-back auditif et visuel (un masque local recouvre brièvement l'objet sélectionné). Le nombre d'objets pertinents présents dans les scènes étant variable selon les essais, la taille du masque local est parfois ajustée, de sorte à englober avec l'objet sélectionné les objets pertinents situés dans son environnement proche. Ainsi, bien que chaque scène visuelle contienne un nombre différent d'objets pertinents, le nombre de clicks nécessaire pour sélectionner les objets pertinents reste constant et compris entre 2 et 4 clicks. Une fois que le participant a sélectionné tous les objets pertinents, il clique sur un bouton « *Terminé* » situé en bas à droite de l'écran. Ce click déclenche l'apparition d'un masque visuel d'une durée de 500ms, dont la fonction est double : (i) masquer le bref mouvement inhérent à l'introduction du changement dans la scène par une interruption visuelle, de sorte à court-circuiter l'orientation exogène de l'attention dans la région du changement (O'Regan, 2001) et (ii) empêcher la persistance rétinienne du stimulus. La présentation de la scène post-changement fait immédiatement suite à la disparition du masque visuel : la tâche du participant est d'indiquer si la scène post-changement présentée est identique à la scène pré-changement (le participant clique sur un bouton « *Pareil* ») ou non (le participant clique sur un bouton « *Pas pareil* »). S'il signale avoir perçu un changement, le participant doit, de surcroît, le localiser (*i.e.*, préciser où se

⁴¹ Cette catégorie est de niveau de base. Tversky & Hemenway (1983; 1984) ont, en effet, montré que les scènes naturelles et les objets qu'elles contiennent peuvent être définis selon trois niveaux de catégorisation : un niveau de base (ex : « voiture »), un niveau super-ordonné (ex : « véhicule ») et un niveau supra-ordonné (ex : « Renault Clio ») et que le niveau de base constitue le niveau de description préférentiellement adopté par les participants dans des tâches de compréhension de scène et de dénomination d'objets.

situe le changement de la scène) et l'identifier oralement (*i.e.*, préciser de quel type de changement il s'agit). Les réponses des participants sont notées par l'expérimentateur. La scène post-changement reste affichée sur l'écran jusqu'à réponse du participant. Les participants montrant différents degrés de familiarité et de dextérité avec l'outil informatique et le maniement de la souris, nous avons décidé, dans cette expérience, de laisser libres les durées de présentation des scènes pré- et post-changements. Enfin, par un click sur un bouton fléché, le participant déclenche l'apparition de l'écran de transition, et un nouvel essai commence. Il en est de même pour tous les essais de l'expérience.

L'expérience se déroule de façon similaire pour les participants du groupe « vision passive », si ce n'est que, confrontés à la scène pré-changement, ces derniers reçoivent simplement comme consigne de regarder la scène, et de cliquer sur un bouton « Terminé » lorsqu'ils estiment avoir regardé la scène suffisamment longtemps.

Préalablement à l'expérience, tous les participants passent une phase d'entraînement sur cinq essais, phase destinée à les familiariser avec la tâche. Les cinq essais se répartissent comme suit : un essai dans lequel les scènes pré- et post-changement sont identiques, et quatre essais dans lesquels les scènes pré- et post-changement diffèrent par l'addition ou par la disparition d'un objet de la scène. Ces quatre derniers essais sont destinés à présenter aux participants les quatre combinaisons de changements possibles (CP, CM, NPC, et NPM).

Au total, la passation de l'expérience requiert trente-cinq minutes environ.

1.2. Analyse des résultats

Les performances des participants sont analysées en termes de score de détection des changements et de latences de réponse. Le score de détection est déterminé selon un critère strict puisque, pour que leurs réponses soient considérées comme correctes, les participants

doivent avoir localisé *et* identifié correctement le changement. Toute autre réponse est considérée comme une erreur. La consigne de l'expérience privilégiant la justesse plutôt que la rapidité de la réponse, l'analyse des résultats considèrera le score de détection comme la mesure principale de la représentation des informations visuelles en mémoire, et les latences des réponses comme un indicateur d'ordre secondaire dans l'interprétation des processus cognitifs sous-tendant les performances des participants. Enfin, dans le groupe «vision active», l'analyse des latences de réponses sera conduite sur les performances des 16 participants, alors que l'analyse des scores détections correctes ne sera conduite que sur les performances de 13 participants seulement⁴². Les données recueillies ont été analysées au moyen d'analyses de la variance (ANOVAs à mesures répétées). Étant donné que les items ne sont pas uniformément répartis dans les conditions expérimentales, nous avons conduit des analyses séparées pour l'effet des facteurs pertinence et intérêt sémantique du changement sur la détection des changements.

Plan d'analyse

$$\underline{S} < G_2 > * P_2$$

$$\underline{S} < G_2 > * I_2$$

Le facteur \underline{S} correspond aux participants ; le facteur G correspond au groupe (Vision active vs. Vision passive) ; le facteur P correspond à la pertinence du changement par rapport à la tâche (Pertinent vs. Non pertinent)⁴³ ; le facteur I correspond à la l'intérêt sémantique du changement (Central vs. Marginal).

⁴² Trois des participants n'ont pas reçu la consigne de localiser et d'identifier oralement le changement détecté.

⁴³ Il n'est, bien entendu, jamais question de pertinence du changement pour le groupe « vision passive ». En fait, la dénomination « changements pertinents » et « changement non pertinents » pour le groupe « vision passive » désigne les changements qui sont pertinents et non pertinents pour le groupe «vision active». Nous conserverons cette dénomination car elle facilite les comparaisons inter groupes et permet, de fait, de déterminer dans quelle mesure la détection d'un même changement dépend de son implication dans la tâche, c'est-à-dire dans quelle mesure l'attribution d'un statut fonctionnel affecte la représentation en mémoire de l'information visuelle.

Les performances obtenues figurent dans le Tableau 1.

Tableau 1 : Scores de détections correctes (%) et latences de réponses (s) en fonction du type de changement et du groupe pour l'expérience 1. VA : groupe «vision active» ; VP : groupe «vision passive» ; P : changement pertinent ; NP : changement non pertinent ; C : changement central ; M : changement marginal. Erreurs type entre parenthèses.

	Pertinence		Intérêt sémantique	
	P	NP	C	M
Scores de détection				
VA	54.95 (4.23)	12.09 (3.91)	30.77 (4.81)	36.26 (3.83)
VP	22.08 (4.02)	35.06 (6.20)	26.66 (5.81)	32.47 (5.48)
Latences de réponses				
VA	7.55 (.45)	10.08 (1.00)	8.79 (.65)	8.84 (.75)
VP	8.88 (.57)	8.92 (.91)	8.67 (.63)	9.13 (.78)

1.2.1. Analyse des scores de détection

Dans un premier temps, l'analyse des scores de détection en fonction de la pertinence du changement révèle un effet de l'interaction du groupe avec la pertinence du changement par rapport à la tâche ($F(1, 22) = 54.27, p < .0001$) : alors que les changements non pertinents sont mieux détectés que les changements pertinents par le groupe «vision passive» ($F(1, 22) = 5.42, p < .05$), le groupe «vision active» montre, a contrario, un meilleur score de détection pour les changements pertinents que pour les changements non pertinents ($F(1, 12) = 69.74, p < .0001$).

Dans un second temps, l'analyse des scores de détection en fonction de l'intérêt sémantique du changement ne révèle aucun effet de l'interaction du groupe avec l'intérêt sémantique du changement ($F(1, 22) < 1, ns.$), pas plus qu'elle ne révèle d'effet principal de l'intérêt sémantique du changement sur les performances de chacun des groupes ($F(1, 22) <$

1, *ns.*, pour le groupe «vision active»; et $F(1, 22) = 1.34$, *ns.*, pour le groupe «vision passive»).

1.2.2. Analyse des latences de réponses

Une première analyse⁴⁴ des latences de réponses comptabilisées pour toutes les réponses (correctes et incorrectes) révèle que les participants des deux groupes répondent plus rapidement aux paires de photographies modifiées qu'aux paires de photographies inchangées. Les latences de réponse moyennes (et les erreurs-type associées) aux paires de photographies changées et inchangées sont respectivement 8.82 (.65) et 9.95 (.68) pour le groupe «vision active» ($t(15) = 2.24$, $p < .05$), et 8.90 (.71) et 11.50 (.1.09) pour le groupe «vision passive» ($t(15) = 3.59$, $p < .005$). Ce résultat est classique dans les tâches de recherche visuelle : les participants répondent plus rapidement à la présence qu'à l'absence d'une cible dans une scène visuelle (Treisman & Gelade, 1980).

Pour les analyses qui suivront, seules les latences de réponses aux changements seront prises en compte. Étant donné que la cécité au changement se caractérise par un défaut de détection des changements, les latences de toutes les réponses aux changements (correctes et incorrectes) seront prises en compte dans l'analyse. Aussi, les analyses de la variance⁴⁵ prendront en considération une co-variable, le nombre de bonnes détections pour chaque type de changement, afin de rendre compte des effets le plus finement possible. Les résultats montrent que le pattern des latences de réponses suit le pattern des scores de détection. Dans un premier temps, les résultats révèlent un effet de l'interaction du groupe avec la pertinence du changement sur les latences de réponses ($F(1, 21) = 4.40$, $p < .05$) : dans le groupe «vision

⁴⁴ Test t de Student sur échantillons appariés

⁴⁵ ANCOVAs à mesures répétées

active», on observe des réponses plus rapides aux changements pertinents qu'aux changements non pertinents ($F(1, 21) = 6.77, p < .05$) ; alors que l'on ne relève pas d'effet de la pertinence du changement sur les latences des réponses dans le groupe «vision passive» ($F(1, 21) < 1, ns$). Dans un second temps, les résultats ne révèlent aucun effet de l'interaction du groupe avec l'intérêt sémantique du changement par rapport à la tâche ($F(1, 21) < 1, ns$), ni d'effet principal de l'intérêt sémantique du changement sur les latences de réponses pour aucun des groupes ($F(1, 21) < 1, ns.$, pour le groupe «vision active» ; et $F(1, 21) < 1, ns.$, pour le groupe «vision passive»).

1.2.3. Analyse des temps d'exploration

Les résultats montrent que les groupes ne se distinguent pas du point de vue des temps d'exploration (21.64 s pour le groupe «vision active» et 19.68 pour le groupe «vision passive» ; $F(1, 25) < 1, ns$).

1.3. Discussion

Dans cette expérience, nous avons analysé isolément les effets respectifs de l'intérêt sémantique des changements et de la pertinence des changements par rapport à la tâche sur la perception de scènes naturelles complexes. Comme il est d'usage dans les expériences de cécité au changement, la détection des changements est utilisée pour évaluer ce qui a été focalisé par l'attention et représenté en mémoire (Shapiro, 2000). L'on peut déjà remarquer que, dans cette expérience, l'envergure de la cécité au changement est de taille, puisque les performances ne dépassent jamais 55% de détections. Une analyse plus détaillée des résultats révèle, en ce qui concerne le groupe «vision active», que les changements sont

significativement mieux perçus lorsqu'ils sont pertinents par rapport à la tâche que lorsqu'ils sont non pertinents, autant du point de vue de la quantité (ils sont perçus en plus grand nombre) que de la rapidité (ils sont perçus plus rapidement). Ce résultat est d'autant plus puissant qu'il contraste significativement avec le pattern de détections observé dans le groupe «vision passive» où ces mêmes changements sont dépossédés de leur statut de pertinence. Les résultats du groupe «vision passive» montrent en effet que les changements les plus à même d'être détectés par les participants sont les changements non pertinents. Ce résultat plutôt surprenant⁴⁶ révèle néanmoins une caractéristique du matériel utilisé dans l'expérience : il semblerait que les changements non pertinents soient perceptivement plus saillants que les changements pertinents, et qu'ils aient été, par conséquent, plus facilement détectés par les participants. Ce dernier point renforce l'effet de la tâche sur les traitements visuels observé dans le groupe «vision active»: la tâche de recherche visuelle mise en œuvre lors de la phase d'exploration semble avoir totalement redessiné les patterns de traitements visuels tels qu'ils auraient été spontanément mis en œuvre par les observateurs. Tout se passe comme si la tâche de recherche visuelle avait littéralement guidé, en phase d'exploration des scènes pré-changement, le déploiement de l'attention focalisée sur les éléments de la scène impliqués dans la résolution de la tâche. En conséquence de quoi, les éléments pertinents par rapport à la tâche auraient été préférentiellement extraits de la scène visuelle, traités et représentés en mémoire de manière suffisamment détaillée pour permettre ultérieurement la détection de leur changement. Ainsi, les buts de la tâche auraient non seulement favorisé le traitement en détail de certains objets de la scène visuelle ainsi que leur représentation en mémoire, mais ils auraient également conduit à préserver ces informations suffisamment longtemps en mémoire

⁴⁶ Il serait légitime d'être surpris par un tel résultat puisque, le groupe contrôle recevant simplement comme consigne de regarder les scènes pré-changement, il ne peut être question d'un quelconque statut de pertinence pour ce groupe. Nous attribuons ce résultat à un biais de matériel, très probablement lié à la complexité des scènes visuelles utilisées dans l'expérience.

pour soutenir la détection ultérieure de leur changement. Avoir été représentée en mémoire conséquemment à la résolution de la tâche qui l'aurait mise en jeu confèrerait donc à l'information visuelle une garantie représentationnelle : elle perdurerait quelque secondes en mémoire, même après que l'attention se serait détournée d'elle. En résumé, dès qu'une tâche serait à réaliser sur une scène visuelle, les informations visuelles pertinentes, rendues saillantes par les exigences de la tâche, bénéficieraient d'une focalisation attentionnelle privilégiée. Cette focalisation attentionnelle entraînerait, dans un premier temps, le traitement approfondi de l'information visuelle directement nécessaire et pertinente par rapport aux exigences de la tâche, et garantirait, dans un second temps, la représentation soutenue de cette information en mémoire, pendant la durée de la tâche.

Bien que les groupes « vision active » et « vision passive » ne se distinguent pas du point de vue des temps d'exploration (ce qui semble suggérer que l'effet facilitateur de la pertinence par rapport à la tâche sur la représentation en mémoire ne s'explique pas par une mémorisation plus efficace des scènes pré-changement, du moins en ce qui concerne les différences de performances observées entre le groupe «vision passive» et le groupe «vision active»), il reste possible que nos résultats reflètent néanmoins les durées d'exploration des scènes pré-changement. Il reste par exemple envisageable que le groupe «vision active» ait exploré plus longuement les scènes pré-changement pertinents que les scènes pré-changement non pertinents, ce qui pourrait par conséquent expliquer les différences de performances observées pour ces deux modalités, et rendrait ipso facto caduque toute influence de la tâche sur la représentation de l'information visuelle en mémoire. De la même façon, les performances affichées par les participants du groupe «vision passive» pourraient s'expliquer, non pas par un biais de matériel comme nous en avons fait l'hypothèse, mais tout simplement par une exploration plus longue des scènes pré-changement non pertinents. Afin de tester cette hypothèse, nous avons analysé les éventuelles relations de corrélation existant entre les

scores de détection des changements pertinents et non pertinents et le temps passé à explorer les scènes pré-changement correspondantes, par les participants des groupes « vision active » et « vision passive ». Si les performances de détection affichées par chacun des groupes résultent d'une meilleure mémorisation des scènes pré-changement, alors l'on devrait observer chez le groupe «vision active» des temps d'exploration des scènes pré-changement pertinents corrélés aux scores de détection des changements pertinents ; et chez le groupe «vision passive», des temps d'exploration des scènes pré-changement non pertinents corrélés aux scores de détection des changements non pertinents. Les résultats infirment ces hypothèses : les temps d'exploration des scènes pré-changement et le scores de bonnes détections ne sont corrélés ni chez le groupe «vision active» ($r_{\text{pertinent}} = .13, ns$; $r_{\text{non pertinent}} = .11, ns$) ni chez le groupe «vision passive» ($r_{\text{pertinent}} = .22, ns$; $r_{\text{non pertinent}} = .35, ns$). Les résultats obtenus ne s'expliquent donc pas par une meilleure mémorisation de tel ou tel stimulus, mais bel et bien par le traitement approfondi des informations visuelles pertinentes par rapport à la tâche.

Alors que les résultats obtenus révèlent sans conteste le rôle majeur de la tâche sur la représentation de l'information visuelle en mémoire, le rôle de l'intérêt sémantique de l'information sur les représentations visuelles reste pour le moins discutable. En effet, nos résultats ne montrent aucun effet de l'intérêt sémantique sur la détection des changements : les changements centraux ne sont pas mieux détectés que les changements marginaux, ni du point de vue des scores de détection, ni du point de vue des latences de réponses, contrairement aux résultats classiquement obtenus dans les expériences de cécité au changement (O'Regan, 2001; O'Regan et al., 2000; Rensink, 2000a; Rensink et al., 1997; Simons & Levin, 1997). Cependant, un test de t de Student⁴⁷ effectué sur les latences de

⁴⁷ Test unilatéral sur échantillon apparié

réponses de sept participants (différents de ceux de l'expérience) soumis à un test de contrôle de la visibilité des changements⁴⁸ montre que les changements centraux sont plus rapidement détectés que les changements marginaux ($t(6) = 1.99, p < .05$) : l'absence d'effet de l'intérêt sémantique sur la détection des changements par les participants de l'expérience ne relève donc pas d'un effet de matériel. Comment peut-on alors expliquer l'absence de l'effet de l'intérêt sémantique sur la détection de changement dans notre expérience ? Plusieurs explications peuvent être avancées. Il est tout d'abord possible que l'effet de l'intérêt sémantique sur la détection des changements n'existe qu'en temps réel, c'est-à-dire tant que l'attention est focalisée sur la zone du changement (Rensink, 2000a; Wolfe, 1999), et ne serait donc observable qu'à l'aide de paradigmes de cécité au changement bien spécifiques, comme le paradigme « flicker » ou le paradigme « mudsplash ». Il est également possible que cette expérience ne mette en scène que trop peu de stimuli pour que l'effet de l'intérêt sémantique du changement sur les détections de changement puisse ressortir. Le fait que ce même nombre de stimuli soit suffisant pour permettre aux effets du facteur « pertinence du changement » sur les performances de détection de se manifester suggérerait néanmoins que ces derniers effets soient plus robustes que les effets du facteur intérêt sémantique du changement. Enfin, il se pourrait aussi que l'effet de l'intérêt sémantique des changements ait été annihilé par le fait que le groupe « vision passive » se soit lui-même assigné une tâche de recherche active des changements : en effet, les participants sachant d'un côté qu'un changement est introduit dans les scènes visuelles et qu'il leur faut le détecter, et d'un autre côté ayant conscience de la difficulté de la tâche, il est possible qu'ils aient adopté en conséquence une stratégie de type

⁴⁸ Ce test de contrôle de la visibilité des changements consiste en la présentation simultanée des scènes pré- et post-changement : la tâche des participants est de détecter, localiser et identifier les changements le plus rapidement possible.

« scruter minutieusement la scène » et/ou « rechercher activement les changements »⁴⁹. La mise en application de cette stratégie produirait un effet similaire à celui produit par la tâche assignée au groupe «vision active», et aurait comme conséquence directe l’annihilation des effets potentiels de l’intérêt sémantique des changements : les changements centraux ne seraient pas mieux perçus que les changements marginaux. Aussi, on peut légitimement supposer que la réduction des durées d’exploration des scènes ainsi que l’accent mis sur l’importance de ne pas mémoriser les scènes visuelles pourraient inciter les participants à abandonner cette stratégie. Un des objectifs de l’expérience 2 était de tester cette nouvelle hypothèse.

2. EXPÉRIENCE 2 : EFFET DES FACTEURS SÉMANTIQUES ET FONCTIONNELS SUR LA PERCEPTION DE SCÈNES NATURELLES COMPLEXES

L’expérience 1 a permis de mettre en évidence l’influence prépondérante de la tâche sur les représentations visuelles : les buts de la tâche guident le déploiement attentionnel sur les éléments de la scène directement nécessaires à la résolution de la tâche, et entraînent par conséquent la représentation de ces éléments en mémoire, de façon suffisamment détaillée pour soutenir la détection de leur changement. Malheureusement, l’expérience 1 ne révèle aucun effet de l’intérêt sémantique de l’information sur les représentations visuelles, et ne

⁴⁹Cette hypothèse est renforcée par l’étude des erreurs-types, qui révèlent que les données du groupe « vision passive » se distribuent de manière beaucoup plus hétérogène que les données du groupe «vision active». Cette hétérogénéité différentielle des données pourrait refléter la dispersion des participants du groupe « vision passive » dans diverses tâches auto-assignées, et le regroupement des participants du groupe «vision active» autour d’une même tâche de recherche visuelle.

permet pas non plus de tester un éventuel effet de l'interaction des facteurs relatifs à la pertinence par rapport à la tâche et à l'intérêt sémantique de l'information sur les représentations visuelles. Cette absence d'effet pouvant résulter soit d'un effet uniquement observable en temps réel, soit de la mise en place stratégique d'un examen minutieux des scènes visuelles en vue du test de détection de changements, un des objectifs de l'expérience 2 était de décourager l'utilisation d'une stratégie de scrutation des scènes visuelles tout en conservant un paradigme de cécité au changement en cycle unique, qui permet d'observer les effets différés des facteurs manipulés sur la détection des changements. Pour ce faire, nous avons donc limité les durées d'exploration des scènes et explicitement insisté auprès des participants sur le fait qu'il ne s'agissait pas d'un test de mémoire mais d'un test de perception, et que par conséquent il était inutile d'essayer de mémoriser les scènes présentées. Enfin, cette absence d'effet pouvant également résulter d'un nombre trop faible de stimuli pour permettre aux effets d'apparaître significativement, le nombre de stimuli a été augmenté dans l'expérience 2.

Dans un dernier temps, et afin de nous permettre d'observer avec finesse l'étendue des effets des facteurs sémantiques et fonctionnels sur la cognition des scènes naturelles, nous avons mesuré à la fois la détection explicite et la détection implicite des changements. Étant donné qu'il existe différents niveaux de traitement de l'information visuelle dans le cerveau (pour une revue, voir Maunsell (1995)), et par conséquent plusieurs niveaux de représentations des scènes visuelles (VanRullen & Koch, 2003; VanRullen & Thorpe, 2001b), la mesure conjointe des détections explicites et implicites des changements devrait nous permettre d'étudier non seulement le contenu des représentations visuelles mais également leur degré d'accessibilité à la conscience (Simons et al., 2003).

2.1. Méthode

2.1.1. Participants

Quatre-vingt-trois étudiants inscrits en licence de Psychologie de l'Université de Provence ont participé à cette expérience. Quarante-huit d'entre eux ont été aléatoirement répartis en deux groupes selon la condition de passation de l'expérience : groupe « vision active » ($n = 24$) et groupe « vision passive » ($n = 24$). Les autres participants ont contribué à contrôler l'intérêt sémantique du matériel ($n = 25$) et la visibilité des changements ($n = 10$). Tous les participants, naïfs du point de vue des objectifs de cette recherche, ont attesté d'une acuité visuelle normale ou corrigée. Aucun d'entre eux n'a participé à l'expérience 1.

2.1.2. Matériel

Cinquante deux photographies couleurs de scènes naturelles complexes (dont quatre utilisées dans la phase d'entraînement) ont été numérisées au format 342 x 512 ou 512 x 342 pixels. Certaines de ces photographies avaient été utilisées dans l'expérience 1. Toutes les photographies ont été dupliquées et numériquement retouchées afin d'obtenir cinquante-deux paires de photographies ne se distinguant que par la modification (addition, disparition, déplacement, ou changement de couleur) de tout ou partie d'un objet de la scène. Tout comme dans l'expérience 1, les changements effectués résultent de la manipulation orthogonale de deux facteurs : la pertinence du changement par rapport à la tâche (pertinent *vs.* non pertinent) et l'intérêt sémantique du changement (central *vs.* marginal), déterminés de la même façon que dans l'expérience 1. Il en résulte quatre combinaisons de changements : Pertinent Central (PC), Pertinent Marginal (PM), Non Pertinent Central (NPC), Non Pertinent Marginal (NPM). Chaque combinaison de changement comporte 3 additions, 3 disparitions, 3 déplacements et 3

changements de couleur (voir annexe 2 pour une illustration de ces changements). Ces changements sont également construits de façon à se conformer aux règles de Simon et Levin (1998), et le test de visibilité des changements est similaire à celui de l'expérience 1⁵⁰. Le masque expérimental utilisé est le même que dans l'expérience 1.

Le test de choix forcé propose quatre items alignés sur une grille invisible : la cible et trois distracteurs. La cible et les distracteurs sont extraits de la scène pré-changement⁵¹, exception faite des changements de type « addition » dans lesquels la cible, absente des scènes pré-changement, est extraite des scènes post-changement⁵². Le test de choix forcé propose systématiquement un item par combinaison de changement, dans un ordre immuable : un item PC, un item PM, un item NPC, et un item NPM. La position de la cible varie de la première à la quatrième place selon le type de changement de l'essai.

Les quarante huit essais sont répartis en quatre blocs de dix essais et un bloc de huit essais selon un ordre aléatoire contrôlé. La répartition des paires de photographies dans les blocs répond à des critères similaires à ceux de l'expérience 1 : chaque bloc doit au moins comprendre un exemplaire de chaque combinaison de changements (PC, PM, NPC, NPM) et de chaque type de changement (addition, disparition, déplacement et changement de couleur). L'ordre des blocs ainsi que l'ordre des paires de photographies à l'intérieur des blocs restent constants pour tous les participants.

⁵⁰ Là encore, tous les changements étaient aisément détectés lorsque les scènes pré- et post-changement étaient présentées côte à côte.

⁵¹ Plusieurs recherches ont montré que la cécité au changement pouvait résulter d'une limitation à comparer les informations visuelles de la scène pré-changement (représentée en mémoire) avec les informations visuelles de la scène post-changement affichée sur l'écran (Angelone et al., 2003; Landman et al., 2003; Mitroff et al., sous presse; Scott-Brown et al., 2000; Simons et al., 2002). Le test de choix forcé tel que nous le proposons permet de faciliter les comparaisons entre les scènes pré- et post-changement en limitant le nombre de comparaisons à quatre objets de la scène. De plus, en proposant une cible extraite de la scène pré-changement, le test de choix forcé permet de réactiver la représentation de la cible telle que présentée dans la scène pré-changement et, par conséquent, de favoriser les comparaisons et faciliter la tâche de détection implicite.

⁵² Étant donné que le changement ne porte que sur la cible, les distracteurs sont les mêmes dans les scènes pré- et post-changements

2.1.3. *Équipement*

L'équipement est le même que dans l'expérience 1

2.1.4. *Procédure : tâche de détection de changements*

La procédure utilisée dans l'expérience 2 est très proche de celle de l'expérience 1 (voir Figure 8). En plus de la consigne écrite, les participants sont vivement encouragés à ne pas mémoriser les scènes qui leur seront présentées. Par ailleurs, les participants sont informés que chacun des essais contient un changement et que le temps alloué à l'exploration des scènes sera limité⁵³, mais toutefois suffisant pour accomplir la tâche de recherche visuelle⁵⁴. Les participants faisant preuve de différents degrés de familiarité avec l'outil informatique, les temps d'exploration ont été déterminés individuellement. Pour ce faire, nous avons soumis tous les participants à une tâche de recherche visuelle lors de la phase d'entraînement, et avons moyenné les durées qui leur étaient nécessaires pour sélectionner, par un click de souris, tous les éléments désignés par la consigne spécifique des quatre scènes pré-changement de l'entraînement. Le temps individuel moyen passé à explorer activement les scènes pré-changement de l'entraînement constitue la durée maximale de présentation des scènes pré- et post-changements de l'expérience, pour chaque participant⁵⁵. Comme dans l'expérience 1, tout nouvel essai commence avec la présentation d'un écran de transition, puis de la scène pré-changement. Les participants du groupe «vision active» doivent réaliser une

⁵³ Information donnée à tous les participants

⁵⁴ Information donnée aux participants du groupe «vision active» seulement.

⁵⁵ Les scènes pré-changement de la phase d'entraînement requièrent entre 4 et 5 clics de souris pour sélectionner les cibles de la tâche de recherche visuelle, alors que les scènes pré-changement de la phase expérimentale à proprement parler en requièrent seulement 3 ou 4. Ainsi, les durées enregistrées pour chaque participant sont pourvues d'une marge de sécurité.

tâche de recherche visuelle sur la scène pré-changement, alors que les participants du groupe «vision passive» doivent simplement la regarder. La scène pré-changement est remplacée par le masque visuel quand les participants cliquent sur un bouton « *Terminé* » (c'est-à-dire quand ils ont terminé la tâche de recherche visuelle en ce qui concerne les participants du groupe «vision active», et quand ils estiment avoir regardé la scène suffisamment longtemps en ce qui concerne les participants du groupe «vision passive») ou quand le temps maximum alloué à l'exploration des scènes pré-changement est atteint. Le masque visuel dure 500 ms et est aussitôt remplacé par la scène post-changement. La tâche des participants est de signaler tout changement perçu et d'en préciser oralement la nature et la position. La scène post-changement reste sur l'écran jusqu'à réponse du participant ou jusqu'à écoulement du temps imparti⁵⁶. Enfin, les participants sont soumis à un test de choix forcé, dans lequel ils doivent indiquer parmi quatre items proposés, lequel a été modifié. Cette étape se fait en temps libre.

L'expérience 2 dure approximativement 50 minutes.

2.2. Résultats

Dans cette expérience, il s'agit d'évaluer les rôles respectifs et conditionnels des facteurs pertinence du changement par rapport à la tâche et intérêt sémantique du changement, ainsi que leur action conjuguée, sur la détection de changements affectant des scènes naturelles complexes. Les scores de détection des participants sont déterminés de la même manière qu'ils l'ont été dans l'expérience 1, c'est-à-dire en adoptant un critère de détection strict : une réponse est considérée comme une détection correcte quand elle identifie et

⁵⁶ Une procédure similaire a été adoptée dans plusieurs expériences antérieures (O'Regan et al., 2000; Rensink et al., 1997, 2000)

localise correctement le changement. Toute autre réponse est considérée comme une erreur, et une absence de réponse comme une non détection. La consigne de l'expérience privilégiant la justesse plutôt que la rapidité de la réponse, l'analyse des résultats considèrera le score de détection et le nombre d'erreurs comme les mesures principales de la représentation des informations visuelles en mémoire, tandis que les latences de réponses seront considérées comme un indicateur d'ordre secondaire.

Dans un souci de clarté, les analyses que nous rapportons ici porteront dans un premier temps sur les performances au test de détection explicite, et dans un second temps, sur les performances au test de détection implicite.

2.2.1. Analyse des performances de détection au test de détection explicite

Les résultats révèlent, dans un premier temps, un effet du type de changement sur les scores de détections⁵⁷ : d'une manière générale, les disparitions et les additions d'objets sont mieux perçues que les déplacements d'objets ou que le changement de leur couleur, autant par les participants du groupe «vision active» ($F(3, 69) = 61.67, p < .0001$) que par les participants du groupe «vision passive» ($F(3, 69) = 36.73, p < .0001$). Ce résultat est en accord avec des résultats antérieurs à cette recherche, lesquels montrent que les changements perceptivement prégnants (additions et disparitions d'objets) sont plus facilement détectés que les changements plus discrets (déplacements et changements de couleur) (Aginsky & Tarr, 2000; Mandler & Ritchey, 1977; Mondy & Coltheart, 2000). Une étude antérieure ayant par ailleurs montré que l'effet de l'intérêt sémantique se retrouve quel que soit le type de

⁵⁷ Analyses intragroupe. Le plan d'analyse est : $S_{24} * T_4$, où le facteur S désigne les participants et le facteur T le type de changement (addition vs. disparition vs. déplacement vs. changement de couleur).

changement proposé (Rensink et al., 1997), nous confondrons la variable « type de changement » dans les analyses qui suivront.

Une première série d'analyses concerne les performances obtenues au test de détection explicite en fonction de la pertinence du changement et de l'intérêt sémantique du changement. Les performances des participants sont mesurées en terme de score de détection, de nombre d'erreurs et de latence de réponses, et sont analysées au moyen d'analyses de la variance (ANOVAs et ANCOVAs, à mesures répétées).

Plan d'analyse:

$$S_{24} < G_2 > * P_2 * I_2$$

Le facteur S correspond aux participants ; le facteur G correspond au groupe (Vision active vs. Vision passive) ; le facteur P correspond à la pertinence du changement par rapport à la tâche⁵⁸ (Pertinent vs. Non pertinent) ; le facteur I correspond à l'intérêt sémantique du changement (Central vs. Marginal)

Les résultats de l'expérience 2 figurent dans le Tableau 2.

⁵⁸ Il n'est, là encore, pas réellement question de pertinence du changement pour le groupe « vision passive », et cela pour la même raison que dans l'expérience 1 (voir note de bas de page n° 43).

Tableau 2 : Scores moyens de détections correctes (%), d'erreurs (%) et latences moyennes (s) en fonction du type de changement et du groupe pour l'expérience 2. VA : groupe «vision active» ; VP : groupe «vision passive» ; P : changement pertinent ; NP : changement non pertinent ; C : changement central ; M : changement marginal ; Erreurs type entre parenthèses.

		P		NP	
		C	M	C	M
Score de détection					
	VA	47.57 (2.32)	44.10 (3.03)	21.18 (2.07)	11.11 (1.98)
	VP	29.86 (2.35)	16.32 (2.38)	24.31 (2.29)	15.28 (1.76)
Erreurs					
	VA	32.99 (3.38)	42.71 (3.49)	50.00 (3.29)	50.00 (4.26)
	VP	34.03 (3.47)	44.44 (3.24)	39.58 (4.14)	41.32 (4.21)
Latences					
	VA	8.15 (.63)	8.64 (.56)	8.41 (.67)	9.76 (.65)
	VP	22.20 (1.73)	24.68 (1.56)	24.37 (1.56)	23.98 (1.64)

Analyse des scores de détection

L'analyse des scores de détection révèle un effet significatif de l'interaction groupe * intérêt sémantique * pertinence du changement ($F(1, 46) = 5.45, p < .05$) : les participants du groupe «vision active» perçoivent davantage de changements que les participants du groupe «vision passive», uniquement lorsque les changements sont pertinents par rapport à la tâche (une analyse des contrastes révèle : $F(1, 46) = 62.78, p < .0001$), sans que rien ne distingue les performances des deux groupes lorsque les changements sont non pertinents par rapport à la tâche (une analyse des contrastes révèle : $F(1, 46) = 2.53, ns$). Ce résultat suggère que l'exploration active des scènes visuelles ne bénéficie au groupe «vision active» que lorsque les participants sont ultérieurement confrontés à des changements pertinents par rapport à la tâche qui a déterminé leur exploration visuelle de la scène. Lorsque les changements sont non pertinents, le guidage actif de l'attention sur les objets impliqués dans la tâche pour ce groupe

ne semble pas avoir d'influence sur la performance des participants, comparativement aux performances du groupe «vision passive».

Un des objectifs de cette seconde expérience était de déterminer plus finement les relations entretenues par les facteurs d'ordre fonctionnel (pertinence du changement par rapport à la tâche) avec les facteurs d'ordre sémantique (intérêt sémantique du changement) ainsi que leur influence sur la perception de scènes naturelles complexes. Les analyses révèlent que les patterns de performances varient en fonction des groupes : alors que les participants du groupe «vision passive» détectent significativement mieux les changements centraux que les changements marginaux ($F(1, 46) = 26.07, p < .0001$), sans qu'il y ait d'effet de la pertinence du changement ($F(1, 46) = 2.19, ns$) ni d'effet de l'interaction ($F(1, 46) = 1.80, ns$) sur le score de détection chez ce groupe, il en est tout autrement pour les participants du groupe «vision active». On retrouve pour ce groupe, conformément aux résultats de l'expérience 1, un effet hautement significatif du facteur pertinence du changement sur les scores de détections ($F(1, 46) = 177.03, p < .0001$) : les participants du groupe «vision active» perçoivent mieux les changements pertinents que les changements non pertinents. L'on observe par ailleurs dans ce groupe une interaction significative entre les facteurs intérêt sémantique et pertinence des changements ($F(1, 46) = 3.85, p = .05$). Cette interaction se manifeste par le fait que les changements centraux sont mieux perçus que les changements marginaux uniquement quand les changements sont non pertinents par rapport à la tâche ($F(1, 46) = 16.48, p < .0005$), quand les changements sont pertinents, l'on n'observe aucune différence de détection entre les changements centraux et les changements marginaux ($F(1, 46) = 1.30, ns$). Ainsi, les résultats montrent que les participants du groupe «vision active» se comportent de la même manière que ceux du groupe «vision passive» quand les changements sont non pertinents par rapport à la tâche, et différemment quand les changements sont pertinents par rapport à la tâche.

Tout comme dans l'expérience 1, les performances mesurées chez les deux groupes ne sauraient s'expliquer par le simple fait d'une mémorisation des scènes pré-changement : il n'y a, en effet, aucune corrélation entre les durées d'exploration des scènes pré-changement et les score de détections pour aucune des combinaisons de changement, ni pour le groupe «vision active» ($r_{PC} = -.00, ns$; $r_{PM} = .33, ns$; $r_{NPC} = -.09, ns$; $r_{NPM} = -.07, ns$), ni pour le groupe « vision passive » ($r_{PC} = .04, ns$; $r_{PM} = -.06, ns$; $r_{NPC} = -.01, ns$; $r_{NPM} = -.24, ns$).

Analyse des erreurs

L'analyse des erreurs ne révèle qu'un effet significatif de l'interaction (d'ordre deux) entre le groupe et la pertinence du changement ($F(1, 46) = 11.93, p < .005$). Cette interaction se manifeste par un plus grand nombre d'erreurs sur la détection des changements non pertinents par les participants du groupe «vision active» que par les participants du groupe «vision passive» (le pourcentage moyen d'erreurs⁵⁹ sur les changements non pertinents égale 50.00 pour le groupe «vision active», et 40.45 pour le groupe «vision passive»), alors que les deux groupes ne se distinguent pas du point de vue des erreurs quand les changements sont pertinents (le pourcentage moyen d'erreurs⁶⁰ sur les changements pertinents: 37.85 pour le groupe «vision active», et 39.24 pour le groupe «vision passive»). En d'autres termes, lorsque la centration expérimentale sur certains objets de la scène pré-changement se révèle trompeuse (*ie.*, lorsque les changements sont non pertinents), les participants du groupe «vision active» font davantage d'erreurs que les participants du groupe «vision passive», qui n'ont pas été soumis à cette centration trompeuse.

⁵⁹ Ces moyennes confondent l'intérêt sémantique des changements

⁶⁰ Ces moyennes confondent l'intérêt sémantique des changements

Analyse des latences de réponse

Il convient tout d'abord de rappeler que, dans cette expérience (et contrairement à l'expérience 1), les participants n'ont la possibilité de répondre que s'ils ont détecté un changement. S'ils n'ont détecté aucun changement, ils doivent attendre la fin du temps imparti à la phase de décision pour passer à l'essai suivant. Par conséquent, afin d'éviter qu'elles ne reflètent les non détections, les analyses⁶¹ des latences de décision dans cette section ne concerneront que les essais pour lesquels les participants ont répondu (que cette réponse soit correcte ou erronée), et prendront en considération une co-variable : le nombre de détections correctes.

Les résultats montrent une interaction significative entre les groupes, l'intérêt sémantique et la pertinence du changement ($F(1, 45) = 5.57, p < .05$) qui se manifeste par le fait que les participants du groupe «vision passive» ont tendance à répondre plus rapidement aux changements centraux qu'aux changements marginaux surtout quand ils sont « pertinents », contrairement aux participants du groupe «vision active», chez lesquels on ne relève aucune différence dans les latences de réponses quelle que soit la combinaison de changement. Enfin, les résultats montrent surtout un effet hautement significatif du groupe : les participants du groupe «vision active» répondent beaucoup plus rapidement que les participants du groupe «vision passive» (les latences moyennes de réponses sont respectivement: 8.74s et 23.81s, $F(1, 45) = 47.08, p < .0001$).

⁶¹ ANCOVAs à mesures répétées

Analyse des temps d'exploration

Les résultats montrent qu'en moyenne, les participants du groupe «vision passive» explorent les scènes pré-changement plus longuement que les participants du groupe «vision active» (les durées moyennes d'exploration sont respectivement 17.01 s et 11.00 s, $t(46) = 8.04$, $p < .0001$), ce qui écarte la possibilité que nos résultats soient le reflet des durées d'exploration, puisque ce sont les participants qui explorent les scènes le moins longtemps qui affichent les meilleures performances au test de détection explicite des changements.

2.2.2. Analyse des performances au test de détection implicite

Afin de déterminer si, à l'instar des détections explicites, les détections implicites sont dépendantes de la stratégie d'exploration des scènes visuelles, nous nous sommes intéressée aux performances des participants au test de choix forcé. Le recours à une mesure de détection implicite du changement a été motivé par plusieurs arguments expérimentaux montrant que le report explicite des changements pouvait sous-estimer le contenu des représentations visuelles, alors que les mesures implicites de détection des changements (telles que l'enregistrement des mouvements oculaires ou le recours à des tests de choix forcé) permettait davantage d'en rendre compte (Fernandez-Duque & Thornton, 2000; Hollingworth et al., 2001; Thornton & Fernandez-Duque, 2000). Dans notre test de détection implicite (test de choix forcé parmi 4 réponses alternatives), les participants doivent reconnaître un des quatre items proposés comme étant l'item changé. Cependant, il est possible que certaines des reconnaissances correctes soient uniquement le fait du hasard : les participants ont en effet une chance sur quatre de sélectionner la cible par hasard. Classiquement, afin de n'évaluer que les performances non imputables au hasard, les chercheurs ont tendance à tester statistiquement, généralement par le biais d'un test t de Student, l'écart de la performance

mesurée chez les participants avec celle donnée par le hasard (soit, ici, une bonne reconnaissance sur quatre, soit un score de détection implicite de 25%). Or le recours à une statistique de Student est inadapté dans le cas présent car certains participants peuvent se révéler plus sensibles au changement que d'autres (Rensink, 2000b, 2004)⁶², et il serait par conséquent inapproprié de tester dans quelle mesure le groupe plutôt que l'individu se comporte différemment que le présumerait le hasard. Aussi, afin de pouvoir évaluer individuellement dans quelle mesure le nombre de reconnaissances correctes n'est pas imputable au hasard, nous avons utilisé un test de discrimination, qui est un traitement statistique très largement répandu dans la recherche agronomique. Typiquement, dans un test de discrimination, les participants sont confrontés à plusieurs échantillons qui, exception faite de l'un d'eux, proviennent tous d'un seul et même produit : la tâche des participants est de déterminer quel est l'échantillon dissemblable, ou, quel échantillon n'est pas issu du même produit que les autres. Notre test de choix forcé s'apparente à un test de discrimination en ce sens où, dans le test de choix forcé, les participants doivent désigner parmi quatre éléments, lequel n'était pas présent dans la scène post-changement⁶³. Afin de traiter les mesures répétées dans les tests de discrimination, plusieurs auteurs conseillent de corriger le nombre d'observations et le nombre de reconnaissances correctes en se basant sur la dispersion des données (Brockhoff & Schlich, 1998; Schlich, Dacremont, & Brockhoff, 2000). En suivant ces recommandations, nous avons analysé⁶⁴ les performances individuelles au test de choix forcé afin de déterminer si les performances mesurées résultaient ou non du hasard. Les

⁶² Rensink estime qu'environ 30% de la population est capable de « sentir » que des changements ont été réalisés sur des scènes visuelles sans pour autant en faire l'expérience consciente, c'est-à-dire être capable de détections explicites.

⁶³ Ou dans la scène pré-changement dans le cas des additions

⁶⁴ Analyse au moyen d'une macro SAS® conçue pour traiter les mesures répétées dans les tests de discrimination, proposée par Brockhoff et Schlich (1998). Ce traitement statistique, associé à des tables de risque (Schlich, 1993), est disponible sur demande à l'adresse électronique schlich@arome.dijon.inra.fr.

reconnaisances correctes utilisées dans cette analyse ont été mesurées uniquement pour les essais pour lesquels les participants avaient échoué au test de détection explicite (par une non détection ou une détection erronée). Aussi, pour pallier le fait que chaque participant comptabilisait un nombre différent d'essais effectifs au test de détection implicite, nous avons rapporté, pour chaque participant, ce nombre de reconnaissances correctes au nombre total de reconnaissances.

Les résultats montrent que 33% des participants (soit 16 participants sur 48) ont réussi le test de choix forcé mieux que s'ils avaient répondu au hasard (avec un risque alpha inférieur à .05, et une puissance de test supérieure ou égale à .82), ce qui montre, conformément aux travaux de Rensink (2004), qu'environ un tiers de nos participants « sent » les changements (« visual sensing ») sans avoir été pour autant capable de les rapporter explicitement au test de détection explicite. Plus spécifiquement, afin de déterminer si la stratégie d'exploration des scènes pré-changement pouvait influencer cette détection implicite des changements et par conséquent influencer l'accès aux représentations visuelles des scènes, nous avons analysé la distribution des détections implicites en fonction du groupe. Les résultats montrent, à l'image des détections explicites, que le pattern des détections implicites se distribue de manière différentielle selon le groupe, puisque 45.83% des participants du groupe «vision passive» (soit 11 participants sur 24) sont capables de détection implicites, alors que seulement 20.83% des participants du groupe «vision active» le sont (soit 5 participants sur 24). Un test de khi-deux révèle un effet tendanciel du groupe sur la réussite au test de détection implicite ($X^2(1) = 3.38, p = .06$) : les participants tendent à réussir davantage le test de choix forcé quand leur exploration des scènes visuelles n'a pas été dévolue à une tâche de recherche visuelle. Ce résultat suggère que les participants du groupe «vision passive», qui ont exploré passivement les scènes pré-changement, ont davantage accès au contenu implicite des représentations

visuelles que les participants du groupe «vision active» qui ont activement exploré les scènes, par le biais d'une tâche de recherche visuelle.

2.3. Discussion

L'objectif de cette expérience était d'évaluer les rôles respectivement joués par les facteurs d'ordre sémantique et les facteurs d'ordre fonctionnels sur la détection de changements dans une scène visuelle naturelle, en condition de vision active, c'est-à-dire quand l'exploration visuelle est dévolue à la réalisation d'une tâche. En d'autres termes, il s'agissait de déterminer si ces deux types d'informations (intérêt sémantique du changement et pertinence du changement par rapport à la tâche) sont traités et représentés en mémoire quand les participants doivent réaliser une tâche de recherche visuelle sur les scènes.

Conformément aux résultats de l'expérience 1, on retrouve un effet hautement significatif de la pertinence du changement sur les scores de détections du groupe «vision active» : les participants perçoivent davantage les changements des objets visuels impliqués dans la tâche de recherche visuelle que les changements des autres objets visuels. Le fait que l'on ne retrouve pas ces effets dans le groupe «vision passive», qui ne doit réaliser aucune tâche de recherche visuelle sur les scènes, révèle que l'effet observé dans le groupe «vision active» ne relève pas de la saillance perceptive des changements⁶⁵, mais résulte bel et bien de la centration active sur les objets visuels impliqués dans la tâche. L'expérience 2 confirme donc ce que l'expérience 1 avait laissé entrevoir : lors de l'exploration visuelle des scènes pré-changement, les buts de la tâche ont préférentiellement guidé l'attention visuelle vers les objets de la scène qui étaient directement concernés par la tâche de recherche visuelle.

⁶⁵ Les changements pertinents seraient perceptivement plus saillants que les changements non pertinents

Focalisés par l'attention, ces objets ont été extraits de la scène visuelle et représentés en mémoire de façon suffisamment détaillée pour soutenir la détection consciente de leur changement, et de façon suffisamment robuste pour que cette représentation perdure en mémoire après que les objets ont été focalisés par l'attention⁶⁶. Tout se passe comme si les buts de la tâche avaient favorisé le traitement approfondi des objets visuels impliqués dans la réalisation de la tâche, autant du point de vue de la qualité de leur représentation que du point de vue de la durée de leur maintien en mémoire.

Contrairement à l'expérience 1, l'expérience 2 permet à l'effet de l'intérêt sémantique du changement sur les représentations visuelles de la scène (typiquement observé dans la littérature, voir Kelley et al., 2003; O'Regan et al., 2000; Rensink et al., 1997, 2000) de se manifester. Ainsi, l'on retrouve chez les participants du groupe «vision passive» un effet de l'intérêt sémantique des changements sur les scores de détection : le changement des objets d'intérêt central, c'est-à-dire des objets importants pour la compréhension et l'interprétation de la scène, est majoritairement mieux perçu que le changement des objets d'intérêt marginal, lesquels n'interviennent que peu, voire pas du tout, dans la compréhension et l'interprétation de la scène. Ainsi, il semblerait que les objets d'intérêt central aient constitué, à un moment du traitement visuel de la scène, une cible préférentielle pour le déploiement de l'œil et de l'attention visuelle, et aient été par conséquent mieux traités et représentés en mémoire que n'ont pu l'être les objets d'intérêt marginal, qui semblent avoir été en très grande partie ignorés du traitement visuel. Plus intéressant encore, les résultats montrent que l'on retrouve aussi cet effet de l'intérêt sémantique du changement sur les performances du groupe «vision active» quand les changements sont non pertinents. Il semblerait donc qu'à un moment donné du traitement visuel de la scène, l'attention se soit focalisée sur les objets d'intérêt central

⁶⁶ En effet, le délai séparant l'exploration des scènes pré-changement du test de détection explicite est au moins de 500 ms (c'est-à-dire la durée du masque expérimental) et peut être relativement long (de l'ordre de plusieurs secondes) selon les durées individuelles maximum.

alors même que l'exploration visuelle des scènes pré-changement était dévolue à la réalisation de la tâche de recherche visuelle. En conséquence, cette focalisation attentionnelle aurait entraîné la représentation en mémoire de ces objets d'intérêt central. Un tel résultat entre en désaccord avec les prédictions de certains modèles computationnels de l'attention visuelle (Navalpakkam & Itti, 2002, 2005), qui postulent que les connaissances stockées en mémoire en long terme et relatives à la scène visuelle (telles que le résumé de la scène, son tracé et son schéma) sont porteuses de suffisamment d'information pour guider, dès les premières étapes du traitement visuel, le déploiement attentionnel et oculaire sur les objets de la scène pertinents par rapport à la tâche. Ces modèles prédisent en effet que, dès la présentation de la scène, l'attention visuelle se focaliserait principalement sur les éléments de la scène pertinents par rapport à la tâche en cours de résolution⁶⁷, sans faire allusion à une quelconque focalisation attentionnelle sur les éléments d'intérêt central de la scène. On peut donc supposer que ces modèles prédisent que les objets d'intérêt central ne devraient pas être suffisamment traités et représentés en mémoire pour que leurs changements soient perçus, puisque, rappelons-le, la focalisation attentionnelle est un élément indispensable à la représentation en mémoire de l'information visuelle et à la détection de ce changement (Hollingworth et al., 2001; Noë & O'Regan, 2000; Rensink, 2001; Rensink et al., 1997, 2000). Ainsi, ces modèles sont dans l'incapacité d'expliquer nos résultats concernant la supériorité de détection des changements centraux sur les changements marginaux par les participants du groupe «vision active», qui doivent réaliser une tâche de recherche visuelle sur les scènes, à moins d'intégrer le traitement précoce de l'intérêt sémantique des objets de la scène, comme le proposent certains modèles théoriques et computationnels (Henderson et al., 1999; Torralba, 2003). Ces derniers modèles postulent que très tôt lors des traitements visuels, les

⁶⁷ Rappelons que dans la majorité des modèles computationnels de l'attention visuelle, le déploiement attentionnel est mesuré via les mouvements oculaires

connaissances contextuelles guident le déploiement de l'attention visuelle sur les objets de la scène qui sont sémantiquement importants, ce qui pourrait, par conséquent, conduire à l'émergence d'un niveau de représentation « sémantique » de la scène visuelle. Ce niveau de représentation sémantique serait rapidement accessible au cours des traitements visuels et cognitifs de la scène, et suffisamment détaillé pour permettre de détecter la présence d'une cible au sein de scènes présentées extrêmement brièvement (Potter, 1976). Nos résultats semblent donc indiquer que les objets d'intérêt central de la scène ont été précocement représentés en mémoire et cela indépendamment du fait que les participants aient eu à traiter activement les scènes visuelles (groupe «vision active») ou non (groupe «vision passive»).

Les résultats de l'expérience 2 montrent donc que la perception d'une scène visuelle complexe, en condition de vision active, relève à la fois de la prise en compte d'informations visuelles de niveau sémantique (*ie.*, intérêt sémantique du changement) et de niveau fonctionnel (*ie.*, pertinence du changement par rapport à la tâche).

Comment ces informations (intérêt sémantique et pertinence par rapport à la tâche) sont-elles représentées en mémoire, et comment vont-elles influencer les comportements? Admettons qu'elles soient codées équitablement au sein d'un même niveau de représentation visuelle. On peut alors supposer que les représentations des objets codés à la fois par leur intérêt sémantique *et* par leur pertinence par rapport à la tâche (*ie.*, les objets PC⁶⁸) devraient être plus saillantes (et/ou plus accessibles) que les représentations des objets seulement codés par leur intérêt sémantique *ou* par leur pertinence par rapport à la tâche (*ie.*, les objets PM⁶⁹ ou NPC⁷⁰). On devrait donc observer de meilleures détections pour le changement des objets doublement codés (PC) que pour les autres changements (PM ou NPC). Or, ce n'est le cas ni

⁶⁸ Pertinent par rapport à la tâche et d'intérêt Central

⁶⁹ Pertinent par rapport à la tâche et d'intérêt Marginal

⁷⁰ Non pertinent par rapport à la tâche et d'intérêt Central

pour les scores de détections ni pour les latences de réponses. Au contraire, on ne trouve aucune différence de détection entre les changements PC et les changements PM, et les seules détections différentiellement observées concernent les objets NPC et NPM⁷¹ (les changements centraux (NPC) sont mieux perçus que les changements marginaux (NPM)). Ce résultat suggère plutôt que les deux types d'informations, intérêt sémantique et pertinence par rapport à la tâche, interviendraient à différents niveaux de traitement et seraient donc codés à différents niveaux de représentation.

Ainsi, considérés ensemble, nos résultats indiquent que la représentation visuelle « sémantique » d'une scène serait créée dès les premières étapes du traitement visuel et serait suivie, lorsqu'une tâche devrait être réalisée sur la base des informations visuelles portées par la scène, par l'émergence d'une représentation visuelle « fonctionnelle », qui fournirait, plus tardivement dans le traitement visuel, un accès détaillé à l'information visuelle immédiatement nécessaire au traitement de la tâche. Ainsi, la capacité des participants du groupe «vision active» à percevoir davantage de changements que les participants du groupe «vision passive» uniquement dans la condition «changements pertinents», tendrait à suggérer que les informations visuelles portées par le plus haut niveau de représentation (*ie.*, le niveau de représentation fonctionnel) seraient plus détaillées et/ou plus accessibles que les informations portées par le niveau de représentation inférieur de la scène (*ie.*, le niveau de représentation sémantique). La tâche conduirait au filtrage et au traitement approfondi des informations visuelles qu'elle implique, et à leur codage au sein d'une représentation visuelle « fonctionnelle », qui serait préférentiellement activée par la tâche et se surimposerait à la représentation visuelle « sémantique » de la scène sans pour autant l'annihiler. Les participants du groupe «vision passive», chez lesquels aucune tâche n'aurait pu conduire à

⁷¹ Non pertinent par rapport à la tâche et d'intérêt Marginal

l'émergence ou l'activation d'un niveau de représentation fonctionnel de la scène, répondraient au test de détection explicite des changements sur la base des informations portées par la représentation « sémantique » de la scène, et par conséquent détecteraient majoritairement les changements opérés sur les objets codés par cette représentation, c'est-à-dire sur les objets d'intérêt central. Au contraire, les participants du groupe «vision active», chez lesquels les buts de la tâche auraient conduit à l'émergence et à l'activation d'un niveau de représentation fonctionnel de la scène, répondraient au test de détection explicite des changements sur la base des informations portées par la représentation « fonctionnelle » de la scène, et par conséquent détecteraient majoritairement les changements opérés sur les objets codés par cette représentation, c'est-à-dire sur les objets pertinents par rapport à la tâche, indépendamment de leur intérêt sémantique. Lorsque les participants du groupe «vision active» auraient à répondre à des changements portant sur des objets de la scène non représentés dans leur représentation fonctionnelle de la scène (c'est-à-dire en condition « changements non pertinents »), ils pourraient néanmoins accéder aux informations visuelles codées par le niveau de représentation « sémantique » de la scène, ce qui expliquerait la similarité des performances observées chez les deux groupes en condition « changement non pertinent ». Ainsi, même si le niveau de représentation de la scène préférentiellement activé par la tâche serait le niveau fonctionnel, le niveau de représentation sémantique, maintenu latent pendant la durée de la tâche, resterait toutefois accessible. Les erreurs de détections observées dans cette expérience pourraient d'ailleurs refléter le coût cognitif représenté par le passage d'un niveau de représentation des scènes à un autre. Les résultats montrent, en effet, que l'on observe un effet de l'interaction groupe * pertinence du changement qui se manifeste par un plus grand nombre d'erreurs de détection sur les changements non pertinents de la part des participants du groupe «vision active» que de la part des participants du groupe «vision passive», alors que les deux groupes ne se distinguent pas du point de vue des erreurs quand

les changements sont pertinents. Ce résultat indique que lorsque la centration expérimentale de l'exploration des scènes pré-changement se révèle trompeuse, les participants du groupe «vision active» font plus d'erreurs que n'en font les participants qui n'ont pas été sujets à cette centration trompeuse, c'est-à-dire les participants du groupe «vision passive». Ce résultat pourrait refléter le coût cognitif engendré par les activités de désengagement des représentations visuelles fonctionnelles et de réactivation des connaissances portées par le niveau de représentation sémantique de la scène, que les participants du groupe «vision active» mettraient en œuvre pour répondre au test de détection explicite.

Les patterns de réponses obtenues au test de détection implicite suggèrent également un accès différentiel aux informations représentées en mémoire selon que les participants ont préalablement reçu la consigne de réaliser une tâche sur les scènes visuelles ou non. Les participants du groupe «vision active» tendent à réussir moins bien au test de détection implicite que les participants du groupe «vision passive». Ainsi, il semblerait que bénéficier de représentations visuelles fonctionnelles faciliterait la détection explicite des changements, mais en gênerait la détection implicite, ce qui suggérerait que la perception implicite des changements serait davantage sous-tendue par les représentations sémantiques que par les représentations visuelles fonctionnelles. Ce résultat est en accord avec une étude antérieure de Mc Conkie et Zola (1979). Dans une expérience, ils ont demandé à des participants de lire un texte présenté en CaSsEs AlTeRnÉeS. Pendant une saccade oculaire, ils changeaient la casse de toutes les lettres (par exemple « LeTtRe » se changeait en « lEtTrE »), ce qui avait pour conséquence de changer littéralement les caractéristiques purement perceptives du matériel sans en modifier ni le sens, ni la syntaxe, ni l'orthographe. Leurs résultats montrent qu'aucune des mesures implicites des changements (ni la durée des fixations oculaires, ni l'envergure des saccades oculaires ni le nombre de mouvements oculaires régressifs) utilisées dans cette expérience n'a pu révéler de perception implicite chez les participants dont la tâche était de

lire le texte : tant que les changements n'affectaient pas les besoins fonctionnels des participants, ils leur demeuraient invisibles. A l'instar de ces résultats, les nôtres semblent indiquer que l'assignation d'une tâche restreint l'accès, conscient ou inconscient, à la représentation de niveau sémantique de la scène.

3. DISCUSSION DU CHAPITRE CINQ

Les expériences présentées dans ce chapitre avaient pour objectif de déterminer dans quelle mesure deux types d'informations visuelles de haut niveau, l'intérêt sémantique des objets de la scène (autrement dit l'importance de ces objets pour la compréhension et l'interprétation de la scène visuelle) et la pertinence de ces objets par rapport à une tâche de recherche visuelle (autrement dit le fait que ces objets aient été ou non la cible de cette tâche), influençaient la détection de changements réalisés sur des scènes naturelles complexes, et par conséquent intervenaient dans la représentation des scènes visuelles en mémoire.

L'expérience 1 réplique l'effet, déjà bien connu, de la pertinence du changement sur la performance de détection (Ballard et al., 1995; Hayhoe, 2000; Hayhoe et al., 1998 ; Hayhoe et al., 2003; Shinoda et al., 2001 ; Triesch et al., 2003; Wallis & Bühlhoff, 2000), sans répliquer l'effet tout aussi bien connu de l'intérêt sémantique du changement sur la performance de détection (Kelley et al., 2003; O'Regan et al., 2000; Rensink et al., 1997, 2000). Concrètement, l'on a pu observer que, lorsque l'exploration visuelle de la scène est dévolue à la réalisation d'une tâche (c'est-à-dire en condition de vision active), les changements des objets de la scène impliqués dans la tâche sont mieux détectés que les changements survenant sur les autres objets visuels, ce qui suggère que les objets de la scène directement impliqués dans la résolution de la tâche sont vraisemblablement mieux représentés en mémoire que les

autres objets de la scène. Par contre, dans cette expérience, nous n'avons pas pu observer d'effet de l'intérêt sémantique des objets visuels sur la détection de leur changement. Or, cette absence d'effet pouvait être imputée à différents biais (nombre de stimuli insuffisant, application d'une stratégie de scrutation appuyée des scènes par les participants), tout comme elle pouvait résulter d'un effet en temps réel, dont l'observation serait incompatible avec le paradigme expérimental utilisé. L'expérience 2 visait donc à permettre l'étude des effets simples et conditionnels des facteurs sémantiques et fonctionnels sur la détection explicite et implicite des changements tout en palliant les possibles biais de l'expérience 1.

Les résultats de l'expérience 2 répliquent en partie les résultats obtenus dans l'expérience 1 : en condition de vision active, les changements pertinents sont mieux détectés que les changements non pertinents, ce qui confirme le rôle joué par les buts de la tâche sur le déploiement de l'attention focalisée au sein de la scène et sur la représentation visuelle qui en découle : les buts de la tâche semblent amener l'attention à se focaliser sur les objets de la scène qui sont directement impliqués dans la réalisation de la tâche et entraîner, par conséquent, le traitement approfondi et la représentation conséquente de ces objets visuels en mémoire de façon suffisamment détaillée pour soutenir la détection consciente de leur changement. Par ailleurs, les résultats de l'expérience 2 révèlent que les représentations visuelles de scènes naturelles ne seraient pas insensibles au facteur d'ordre sémantique, même en condition de vision active, c'est-à-dire quand les participants doivent réaliser une tâche sur les scènes. En effet, dans cette condition, l'effet de l'intérêt sémantique du changement (les changements centraux sont mieux perçus que les changements marginaux) s'observe uniquement lorsque les changements sont non pertinents par rapport à la tâche ; cet effet est alors, dans cette condition, similaire à celui observé en condition de vision passive. Ce premier résultat suggère qu'il existerait plusieurs (hauts) niveaux de représentation d'une

même scène visuelle : un niveau de représentation sémantique de la scène qui coderait les objets permettant la compréhension et l'appréhension de la scène (*ie.*, les objets d'intérêt central), et un niveau de représentation fonctionnel des scènes qui coderait spécifiquement les objets pertinents par rapport à la tâche.

Ainsi, dès que les participants seraient confrontés à une scène visuelle naturelle, le résultat des traitements préattentifs (l'agencement de la scène et la carte de saillance) ainsi que l'activation en mémoire de connaissances sur la scène (l'essentiel et le schéma de la scène) qui interviennent très précocement dans le traitement visuel, concourraient à guider l'attention visuelle au sein de la scène vers les éléments de la scène qui sont déterminants pour en comprendre la signification : les éléments d'intérêt central. Focalisés par l'attention, ces éléments sémantiquement importants seraient codés en mémoire au sein d'une représentation sémantique de haut niveau, qui pourrait également intégrer les connaissances sur la scène stockées en mémoire à long terme (telles que le schéma de la scène et son résumé). Lorsqu'une tâche devrait être réalisée sur la scène, les buts de la tâche guideraient les yeux et la focalisation de l'attention spécifiquement vers les objets de la scène pertinents par rapport à la tâche. Traités en profondeur, ces objets seraient codés au sein d'une représentation fonctionnelle de niveau supérieur, déterminée par la tâche. Ainsi, les représentations visuelles de niveau fonctionnel ne contiendraient que les informations visuelles pertinentes par rapport à la tâche, indépendamment de leur intérêt sémantique. Ce niveau, préférentiellement activé par la tâche à résoudre, pourrait servir la conscience de la situation⁷², et serait maintenu actif (tandis que le niveau sémantique serait maintenu latent) tant que la tâche ne serait pas achevée. Quand l'information codée au sein de la représentation

⁷² Pour davantage d'informations sur la conscience de la situation, voir Endsley (1995)

fonctionnelle ne serait plus suffisante pour répondre au test de détection des changements (*ie.*, pour les changements non pertinents), les participants pourraient réactiver le niveau sémantique de la représentation de la scène, qui serait le niveau activé par les participants placés en condition de vision passive. Ainsi, le niveau sémantique de représentation des scènes constituerait le niveau de représentation de référence lorsque l'exploration des scènes n'a pas été dévolue à la satisfaction d'une tâche, ou que les informations portées par le niveau fonctionnel de représentation s'avèreraient inappropriées. Ce niveau de représentation, activé par défaut lorsque nous serions confrontés à une scène visuelle naturelle, pourrait également servir de cadre de référence à la focalisation de l'attention sur les éléments de la scène qui s'avèreraient pertinents par rapport à une nouvelle tâche.

Notre conception du niveau de représentation sémantique des scènes visuelles complexes partage certaines des caractéristiques des représentations conceptuelles de scènes, qui sont des représentations cohérentes des scènes visuelles (Potter, 1976, 1993, 1999; Potter et al., 2004). Selon Potter, l'information codée au sein de ces représentations serait de nature visuelle (comme nous supposons l'être les informations codées au sein du niveau sémantique des représentations visuelles) et conceptuelle (et pourrait plutôt renvoyer au résumé et au schéma de la scène) ; elle pourrait être extraite de la scène en 173 ms et persister en mémoire pendant au moins 14 secondes (voire plus longtemps). De plus, quand nous supposons que les représentations sémantiques pourraient à leur tour contribuer à enrichir le stock de connaissances sur les scènes en mémoire à long terme (MLT), Potter postule que l'information contenue dans les représentations conceptuelles nécessite d'être consolidée pour être stockée en MLT (Potter et al., 2002). Notre niveau de représentation sémantique des scènes naturelles s'apparente également aux représentations conceptuelles de scènes, notamment en ce qui concerne la rapidité avec laquelle le stimulus est identifié et atteint un niveau de signifiante représentationnelle. Cet accès rapide à la représentation sémantique des

scènes visuelles est également corroboré par des travaux récents (Grill-Spector & Kanwhisher, 2005) qui montrent que la détection d'un objet visuel est concomitante à la catégorisation de niveau basique de cet objet (*eg.*, il s'agit d'un oiseau, d'une ville, d'une voiture ...) et suggèrent par conséquent que des durées d'exposition de scènes naturelles extrêmement brèves (50-150 ms) permettent l'accès au niveau sémantique des objets visuels. Enfin, Potter propose un oubli très rapide de l'information qui n'est pas vouée à être traitée plus en profondeur, ce qui expliquerait que les objets d'intérêt marginal, qui ne sont pas impliqués dans la compréhension de la scène, ne soit pas représentés en mémoire. Par contre elle suppose aussi qu'une des caractéristiques des représentations conceptuelles reposerait sur leur capacité à se restructurer rapidement de manière à pouvoir répondre aux buts fixés par les participants. Plutôt que de supposer une restructuration des représentations visuelles en fonction de la tâche, nos travaux nous conduisent à plutôt parler d'activation préférentielle des informations pertinentes par rapport à la tâche avec émergence d'un niveau fonctionnel de représentation visuelle. Considérés ensemble, les niveaux de représentations sémantique et fonctionnel des scènes visuelles pourraient être opérationnels dès les premiers regards sur les scènes, comme le suggère Potter « *This whole cycle –identification of stimuli, memory recruitment, structuring, consolidation and forgetting of nonstructured material –may occur in less than 1 sec when viewing a pictured scene* » (Potter, 1993, p. 156).

Bien entendu, supposer que la perception d'une scène naturelle serait sous-tendue par deux niveaux de représentations visuelles, l'un codant les informations sémantiques et l'autre codant les informations fonctionnelles portées par la scène peut sembler très coûteux du point de vue des ressources cognitives. Pourtant, l'adhésion à l'idée qu'un même niveau de représentation code pour les deux types d'informations (pertinence par rapport à la tâche et intérêt sémantique) pose, en fait, plus de questions qu'elle ne permet d'en résoudre. En effet,

si un seul niveau de représentation intègre les informations de type sémantique et de type fonctionnel, alors on peut supposer que ces deux types d'informations seront extraites séquentiellement lors du traitement visuel de la scène, comme l'ont montré certains travaux (Hayhoe et al., 2003; Henderson et al., 1999)⁷³. Conformément à ces résultats, l'hypothèse « un même niveau de représentation codant pour les deux types d'informations » suggérerait qu'une scène visuelle est d'abord traitée selon un critère sémantique et ensuite selon un critère fonctionnel. En conséquence, la représentation visuelle d'une scène coderait les objets sémantiquement importants pour l'interprétation de la scène, et ensuite les objets pertinents par rapport à la tâche. Ainsi, avec l'émergence d'une tâche, la représentation sémantique de la scène évoluerait en représentation fonctionnelle. On peut supposer, dans ce cas, que les informations sémantiques seraient soit écrasées par les informations fonctionnelles, soit intégrées, de façon cumulative, aux informations fonctionnelles.

L'alternative « écrasement des informations sémantiques par les informations fonctionnelles » amène à la prédiction que les informations sémantiques, préalablement focalisées par l'attention, seraient effacées par les informations fonctionnelles ultérieurement focalisées (pour des prédictions similaires, voir Rensink, 2000a), et l'on ne devrait pas, par

⁷³ Dans une expérience, Henderson et ses collaborateurs (1999) enregistrent les mouvements oculaires de participants qui doivent déterminer le plus rapidement possible si une cible est présente dans une scène. Cette cible est soit un objet sémantiquement consistant (eg., un mixeur dans une cuisine), soit un objet sémantiquement inconsistant (eg., un microscope dans une cuisine). Les auteurs montrent que la cible est trouvée plus rapidement et fixée plus précocément quand elle est sémantiquement consistante. Ils montrent également que les cibles sémantiquement consistantes sont fixées après une saccade de plus grande amplitude que les cibles sémantiquement inconsistantes. Les résultats de ces travaux vont dans le sens d'une exploration visuelle de scènes naturelles qui répondrait, dans un premier temps, à un critère sémantique puis, dans un second temps, à un critère fonctionnel, puisque, des deux éléments pertinents par rapport à la tâche de recherche visuelle (*ie.*, les cibles), celui qui est trouvé le plus facilement est celui qui est sémantiquement consistant. Les travaux de Hayhoe et collaborateurs (2003) montrent, quant à eux, que lorsque les participants doivent réaliser une tâche, les premières saccades conduisent l'œil sur divers éléments de la scène, qu'ils soient pertinents par rapport à la tâche ou non, alors que les saccades plus tardives sont quasi exclusivement orientées vers les éléments de la scène pertinents par rapport à la tâche. Autrement dit, même quand l'exploration visuelle de la scène est dévolue à la satisfaction d'une tâche de recherche visuelle, les premiers traitements visuel semblent être dirigés vers des éléments qui permettent d'appréhender la scène.

conséquent, observer d'effet de l'intérêt sémantique des objets sur la détection de leur changement. Nos résultats invalident cette alternative.

L'alternative « intégration cumulative des informations sémantiques et fonctionnelles » nous amène à supposer, quant à elle, que les objets de la scène bénéficiant d'un double codage (du fait de leur intérêt sémantique *et* de leur pertinence par rapport à la tâche) devraient être mieux représentés que les objets ne bénéficiant que d'un simple codage (du fait de leur intérêt sémantique *ou* de leur pertinence par rapport à la tâche). Or, nos résultats invalident également cette alternative (voir plus haut dans le texte).

Enfin, l'on pourrait imaginer une dernière alternative de type « intégration pondérée des informations sémantiques et fonctionnelles », qui postulerait que les informations sémantiques et fonctionnelles de la scène seraient traitées et représentées en mémoire avec différents poids d'activation. Cette alternative soulève alors l'épineuse question du devenir de ces informations lorsque la tâche est terminée et que s'avèrent obsolètes les informations représentées en mémoire qui lui étaient associées. Comment le système cognitif parviendrait-il à gérer le désencodage partiel des objets PC, dont la représentation était différenciellement pondérée en mémoire par leur intérêt sémantique et par leur pertinence passée par rapport à la tâche ? Ce problème du désencodage partiel des éléments PC consécutivement à la réalisation de la tâche ne se pose plus si l'on fait l'hypothèse de plusieurs niveaux de représentation de la scène, puisque cette hypothèse suppose, conformément aux conclusions de travaux antérieurs, qu'une fois la tâche achevée, la représentation des informations qui y sont associées décline en mémoire (Ballard et al., 1995; Hayhoe, 2000; Hayhoe et al., 1998).

Afin d'étudier plus en détail le devenir de la représentation des informations pertinentes en mémoire au fur et à mesure que la résolution de la tâche progresse, nous nous sommes plus spécifiquement intéressée, dans les chapitres suivants, au niveau de représentation

fonctionnelle de scènes visuelles complexes. Plus particulièrement, nous avons axé nos travaux sur l'aspect dynamique de ce niveau fonctionnel des représentations visuelles, et avons étudié plus en détail comment il évoluait en fonction du degré d'avancement de la résolution d'une tâche.

– CHAPITRE SIX –

EFFET DE LA RÉOLUTION DE LA TÂCHE SUR LA REPRÉSENTATION DE L'INFORMATION VISUELLE EN MÉMOIRE

Les expériences décrites dans ce chapitre ont pour objectif de déterminer comment, en condition de vision active, les informations de la scène sont codées à un niveau de représentation fonctionnel, et comment le contenu de ces représentations visuelles évolue avec la résolution de la tâche. Pour ce faire, nous avons couplé une tâche de détection de changements avec une tâche de résolution de problème. Nos observations ont porté sur les détections différentielles des changements, en fonction du rang des changements dans la procédure de résolution de problème.

Dans les expériences présentées dans ce chapitre, les participants doivent déplacer, à l'aide de la souris, un véhicule dans le plan d'une ville projeté sur un écran, avec pour but d'arriver le plus rapidement possible à une destination spécifiée. Pour ce faire, ils doivent accomplir un trajet contenant 7 étapes (représentées sur le plan par une icône et un label ; voir Figure 9) dans un ordre précis, la 7^e étape étant l'étape but, c'est-à-dire la destination finale. Ce trajet est déterminé par un arbre des tâches illustré en Figure 10 qui suppose que, pour pouvoir atteindre l'étape but, les participants aient préalablement atteint les deux étapes (A, B) qui lui sont sous-ordonnées. Mais, similairement, pour pouvoir accéder à ces deux étapes, les participants doivent avoir préalablement atteint les 2 étapes respectivement sous-

ordonnées aux étapes A et B (C, D ; et E, F). Cette tâche s'apparente donc à une tâche de résolution de problème en ce sens où les participants doivent, à partir d'un état initial, parvenir à un état but, en respectant un certain nombre de règles qui les contraignent à résoudre plusieurs sous-buts avant de pouvoir atteindre l'état but. A un moment donné de la réalisation du trajet, une étape du plan est changée (son icône est déplacée ou disparaît de la carte)⁷⁴ et le participant a pour consigne de signaler tout changement perçu (c'est-à-dire de préciser quel type de changement affecte quelle étape). Le facteur critique de l'expérience concerne le statut de l'étape modifiée. Cette étape fait partie ou non⁷⁵ de la tâche, c'est-à-dire du trajet à réaliser. Si cette étape fait partie de la tâche, elle peut être impliquée soit dans le sous-but en cours de résolution, soit dans un sous-but antérieurement résolu, soit dans un sous-but ultérieur au sous-but en cours de résolution ou soit au but final. L'expérience permet donc de mesurer la détection des changements par les participants, en fonction du degré d'implication de l'étape changée dans la résolution de la tâche.

Nos prédictions sont plurielles : si, comme l'ont montré Triesch, Ballard, Hayhoe, & Sullivan (2003) les yeux se posent préférentiellement sur les éléments de la scène impliqués dans la réalisation de la tâche, et si nous transférons ce résultat à la mesure explicite de la détection des changements, alors on devrait observer que le système cognitif représente préférentiellement l'information directement utile et pertinente par rapport à la tâche à accomplir. Cela implique que nos représentations visuelles répondraient à un critère de fonctionnalité : le système cognitif ne représenterait que l'information pertinente par rapport à la tâche, avec priorité de représentation donnée aux éléments les plus pertinents à l'instant t .

⁷⁴ Dans l'expérience 3, les changements réalisés sur les scènes sont soit des déplacements soit des disparitions d'étapes, alors que dans les expériences 4 et 5, les changements seront exclusivement des déplacements d'étapes.

⁷⁵ Dans l'expérience 3, l'étape changée fait toujours partie de la tâche, ce qui n'est pas le cas des expériences 4 et 5.

Si tel est le cas, l'on devrait observer qu'une étape a d'autant plus de chances d'être représentée que :

- elle est pertinente par rapport à la tâche
- elle est pertinente par rapport au sous-but en cours de résolution.

Les trois expériences suivantes visent à tester ces hypothèses.

1. EXPÉRIENCE 3 : INFLUENCE DE LA RÉOLUTION DE LA TÂCHE SUR LA DYNAMIQUE DES REPRÉSENTATIONS VISUELLES

L'expérience 3 vise tout particulièrement à déterminer comment le contenu des informations visuelles évolue avec la résolution de la tâche. Nous avons donc étudié la détection de changements d'une scène visuelle complexe quand les changements affectent différentes étapes (ou sous-buts) de la tâche. Nous posons l'hypothèse que les buts de la tâche vont progressivement guider les yeux et l'attention sur les éléments de la scène qui seront directement nécessaires et pertinents pour résoudre le sous-but en cours de résolution. Par conséquent, ces éléments devraient être préférentiellement traités et représentés en mémoire, au sein d'un niveau de représentation fonctionnel de la scène, et leur changement devrait être détecté.

1.1. Méthode

1.1.1. Participants

Soixante-sept étudiants inscrits en licence de Psychologie de l'Université de Provence ont pris part à cette expérience. Ils ont été aléatoirement répartis en deux groupes selon la condition de passation de l'expérience : condition « déplacement » ($n = 41$) et condition « disparition » ($n = 26$). Tous les participants ont attesté d'une acuité visuelle normale ou corrigée et étaient naïfs du point de vue des objectifs de cette recherche. Aucun d'entre eux n'a participé aux expériences 1 et 2.

1.1.2. Matériel

Les cartes

Les stimuli consistent en 6 cartes d'une ville, au format 665 x 508 pixels. Chaque carte contient 13 étapes⁷⁶, chacune d'elle étant représentée sur la carte par une icône et un label. Une carte plus petite contenant quatre étapes⁷⁷ a été créée pour la phase d'entraînement. La Figure 9 donne deux exemples des cartes utilisées dans l'expérience (voir annexe 3 pour une illustration exhaustive des cartes utilisées dans l'expérience).

⁷⁶ Les 13 étapes sont : la gare, la cabine téléphonique, l'entrepôt, le stade, l'église, la préfecture, Paris, l'école, le parc, l'aéroport, la bibliothèque, le kiosque, et le bureau de poste.

⁷⁷ L'opéra, le cinéma, le port et le marché

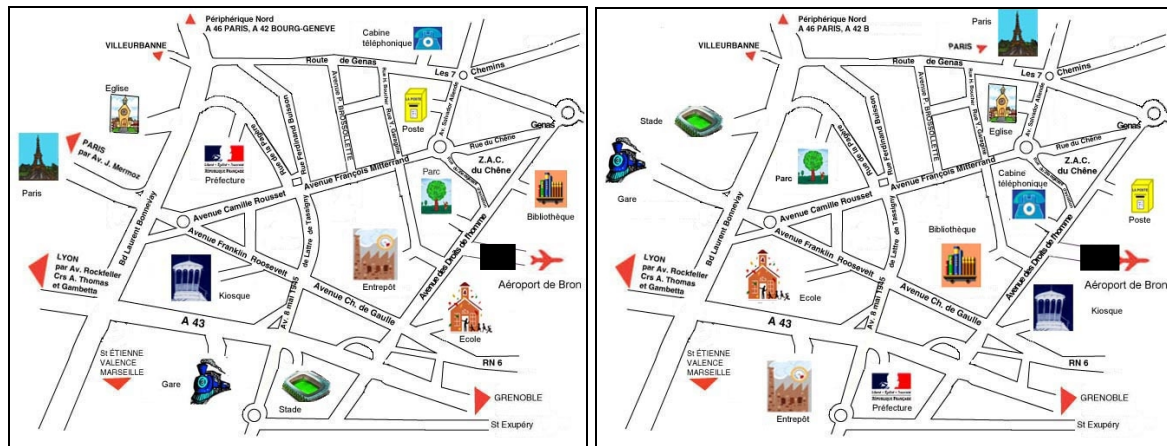


Figure 9 : Exemples de cartes. Le participant doit déplacer un véhicule (représenté par un rectangle noir, ici à l'aéroport) et accomplir un itinéraire en 7 étapes dans un ordre précis, spécifié par l'arbre des tâches illustré en Figure 10. Chaque étape est représentée par une icône et un label. Le changement n'affecte qu'une étape du plan et peut être soit la disparition de l'icône de l'étape, soit le déplacement de l'icône de l'étape dans le plan. Au début de l'expérience, le véhicule est localisé à l'aéroport.

Comme le montre la Figure 9, les 6 cartes sont identiques du point de vue de la configuration routière et diffèrent seulement par la position des étapes dans la carte⁷⁸. La position des étapes dans chaque carte a été déterminée de manière pseudo aléatoire, avec pour contrainte que les icônes des étapes sous-buts proches dans la procédure de résolution de la tâche ne soient pas proches géographiquement. A chacune des cartes correspond le changement de l'une des étapes de la carte, à l'exception d'une carte sur laquelle on procède au changement de six étapes (soit de toutes les étapes mentionnées par l'arbre des tâches, sauf une, celle sur laquelle se trouve le véhicule des participants⁷⁹). Deux types de changements ont été réalisés : disparition d'une étape et déplacement d'une étape dans la carte. Le véhicule que les participants doivent déplacer est représenté par un rectangle noir (voir Figure 9). Le masque visuel est constitué de plusieurs fragments extraits des différents plans, renversés et superposés (voir annexe 3).

L'ordre de passation des six cartes est aléatoire.

⁷⁸ Exception faite de l'icône de départ (l'aéroport) dont la position reste la même dans toutes les cartes.

⁷⁹ Pour une procédure similaire, voir Hayhoe et al. (1998)

L'écran de signalisation des changements

L'écran de signalisation des changements permet aux participants de désigner le changement qu'ils ont perçu sur la scène, en indiquant (i) quel type de changement ils ont perçus et (ii) quelle était l'étape altérée par ce changement. Cet écran propose donc quatre choix possibles sur le type de changement survenu (« image en plus », « image en en moins », « image déplacée » et « image différente »), ainsi la possibilité pour les participants d'indiquer qu'ils ne savent pas de quel type de changement il s'est agi (« je ne sais pas »). Cet écran affiche, en plus, l'icône de chacune des 13 étapes présentes dans le plan, et les participants doivent désigner quelle a été l'étape changée en cliquant sur l'icône correspondant à son choix.

L'arbre des tâches

Le trajet à effectuer dans la carte concerne 7 des 13 étapes contenues dans la carte : la gare, la cabine téléphonique, l'entrepôt, le parc, l'église, la préfecture et Paris. Ce trajet est représenté sur la carte par un arbre des tâches, illustré en Figure 10.

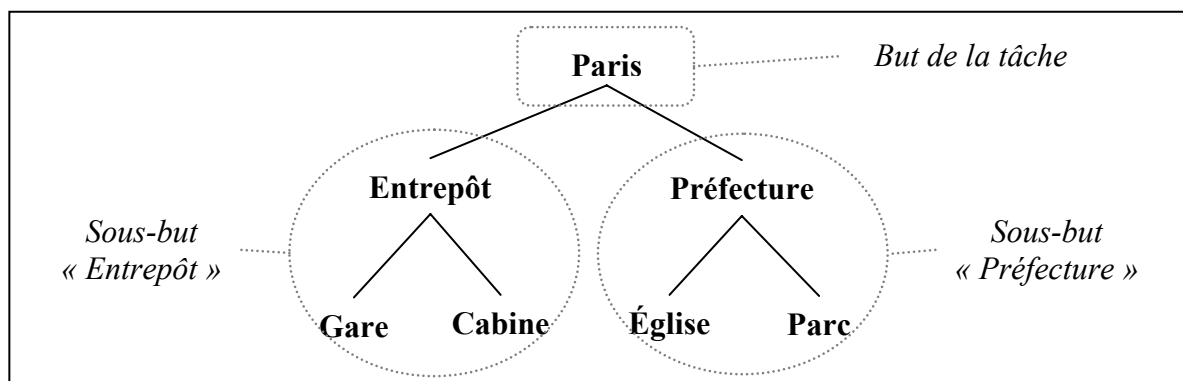


Figure 10 : Arbre des tâches. Il définit l'itinéraire à accomplir. Ici, le but est d'atteindre Paris. Pour cela, les participants doivent avoir atteint deux sous-buts, c'est-à-dire s'être préalablement rendus à l'entrepôt et à la préfecture. Mais ces deux étapes ne peuvent être atteintes qu'après atteinte des deux étapes qui leur sont sous-ordonnées (ie., la gare et la cabine pour l'entrepôt ; l'église et le parc pour la préfecture).

L'arbre des tâches, qui figure toujours en bas à droite de l'écran, définit les conditions de réalisation de la tâche. Comme le montre la Figure 10, les étapes impliquées dans la résolution de la tâche sont organisées en un but (atteindre Paris) et deux sous-buts (atteindre l'entrepôt et la préfecture), qui ne peuvent également être atteints que si les étapes qui leur sont sous-ordonnées ont été préalablement atteintes (la cabine et la gare pour l'entrepôt ; l'église et le parc pour la préfecture). Pour pouvoir réaliser la tâche et atteindre l'étape but (Paris) les participants doivent respecter trois règles :

- Les étapes sous-buts super-ordonnées ne peuvent pas être atteintes si les étapes sous-but qui leur sont sous-ordonnées n'ont pas été préalablement atteintes (par exemple, on ne peut pas se rendre à la préfecture si l'on est pas préalablement passé par l'église et par le parc).
- Les étapes sous-buts de même niveau ont la même priorité : l'on peut commencer par l'une ou l'autre des étapes sous-buts de même niveau (par exemple, l'on peut commencer soit par la branche « entrepôt » soit par la branche « préfecture » ; et au sein de chaque branche, par l'une ou l'autre étape sous-ordonnée - gare ou cabine téléphonique pour l'entrepôt; église ou parc pour la préfecture -)
- La résolution de l'arbre des tâche se fait par branche : il n'est pas possible de rallier d'abord toutes les étapes sous-ordonnées (gare, cabine, parc, église) puis de rallier les étapes super-ordonnées (entrepôt, préfecture) pour arriver à l'étape but.

Par conséquent, pour pouvoir atteindre Paris, il faut avoir résolu les sous-buts « entrepôt » et « préfecture », et chaque sous-but peut être résolu en commençant par atteindre l'une ou l'autre étape sous-ordonnée, ce qui donne huit trajets possibles :

Si on commence par résoudre le sous-but « entrepôt »

- gare → cabine → entrepôt → parc → église → préfecture → Paris
- gare → cabine → entrepôt → église → parc → préfecture → Paris
- cabine → gare → entrepôt → parc → église → préfecture → Paris
- cabine → gare → entrepôt → église → parc → préfecture → Paris

Si on commence par résoudre le sous-but « préfecture »

- église → parc → préfecture → gare → cabine → entrepôt → Paris
- église → parc → préfecture → cabine → gare → entrepôt → Paris
- parc → église → préfecture → gare → cabine → entrepôt → Paris
- parc → église → préfecture → cabine → gare → entrepôt → Paris

L'arbre des tâches reste le même pour les 6 cartes.

Les changements

A un moment donné de la résolution du trajet, un changement (déplacement ou disparition d'étape) survient sur l'une des étapes impliquée dans le trajet (et, dans une condition particulière, sur 6 des étapes du trajet). Ce changement est déclenché par l'arrivée du véhicule des participants sur une étape spécifique. Le facteur crucial de l'expérience est le rang de l'étape changée dans l'ordonnancement des sous-buts de la tâche (ou structure des sous-buts).

Dans l'expérience, le changement survient donc, à chaque essai, sur une des étapes de la carte. Cette étape peut être (i) soit impliquée dans le sous-but en cours de résolution (étape sous-but de **rang n** , lorsqu'elle est immédiatement nécessaire à la résolution du sous-but ; et étape sous-but de **rang $n+1$** , lorsqu'elle suit directement l'étape qui est immédiatement

nécessaire à la résolution du sous-but) ; (ii) soit impliquée dans un sous-but atteint antérieurement (étape sous-but de **rang $n-1$**), (iii) soit impliquée dans un sous-but ultérieur (étape sous-but de **rang $n > 1$**), (iiii) soit impliquée dans le but de la tâche (étape de **rang n_{final}**). Dans toutes ces conditions, le changement ne concerne qu'une étape répertoriée dans l'arbre des tâches.

Enfin, dans une condition un peu particulière, l'on procèdera au changement de toutes les étapes sous-buts, sauf celle sur laquelle le participant se trouve : dans ce cas-là, six changements surviennent simultanément sur la carte. La Figure 11 montre quels changements sont déclenchés par l'atteinte de telle étape, quand le participant commence à parcourir la carte par la branche « entrepôt ».

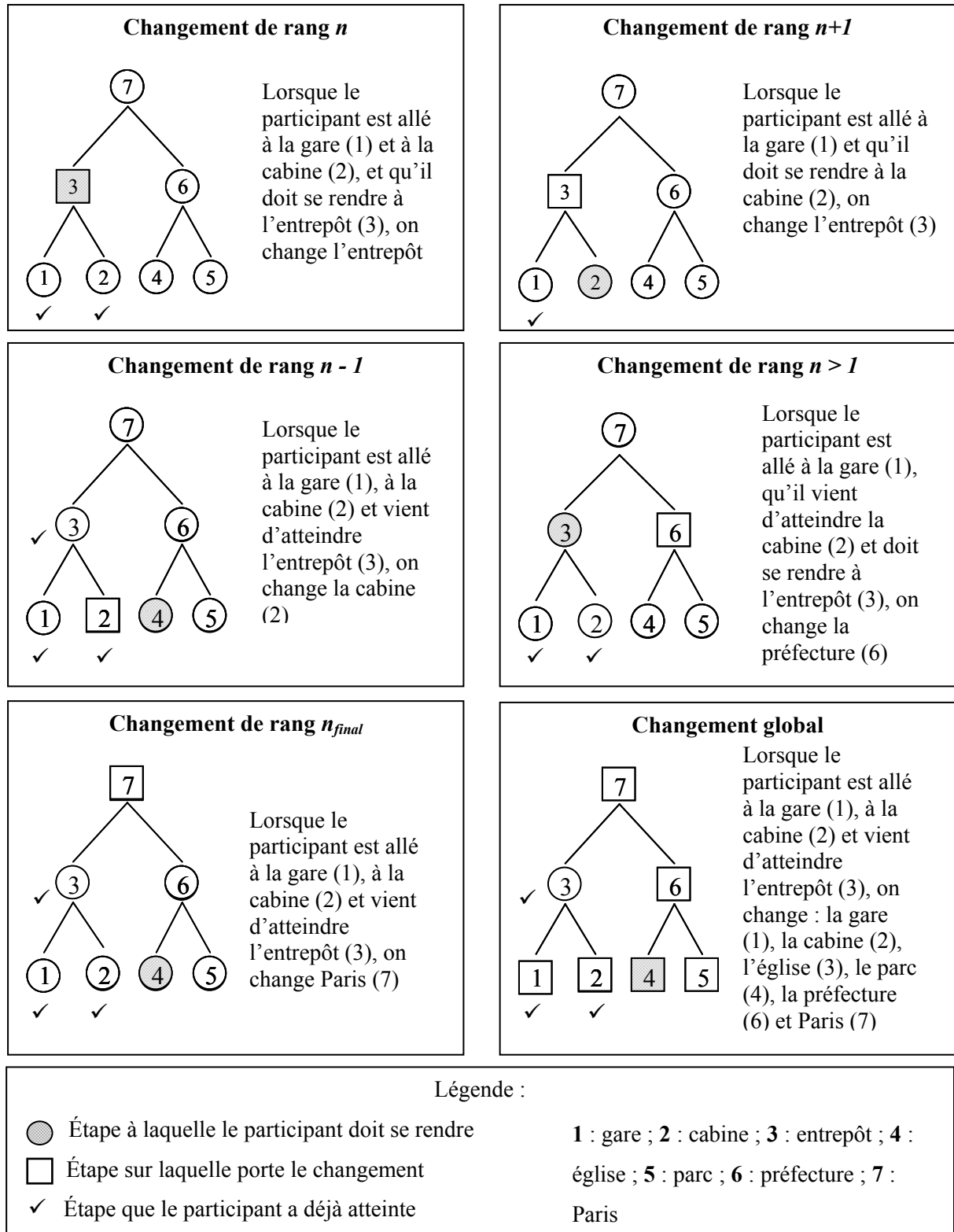


Figure 11. Présentation schématique des différents rangs du changement et du moment de leur survenue sur les cartes. Ce schéma illustre les changements que rencontre un participant effectuant le trajet : gare (1) → cabine (2) → entrepôt (3) → parc (4) → église (5) → préfecture (6) → Paris (7)

Bien entendu, le programme informatique gère de manière correspondante le déclenchement des changements pour les participants qui commencent à parcourir la carte par la branche « préfecture ». Ainsi, quelle que soit la branche par laquelle les participants commencent à parcourir la carte, les changements surviennent au même moment de la résolution de la tâche. Nous faisons remarquer ici que cette dernière indication ne vaut que pour les conditions à « un seul changement ». Pour la condition « six changements simultanés », une limitation du matériel informatique a empêché la programmation du déclenchement des changements en miroir quand les participants commencent par satisfaire la branche « préfecture » de l'arbre des tâches. Aussi, dans cette dernière condition, les six changements sont déclenchés à l'atteinte de l'entrepôt, quelle que soit la branche par laquelle les participants ont commencé leur parcours⁸⁰.

Les changements opérés consistent soit en la disparition de l'icône et du label de l'étape, soit en leur déplacement au sein de la carte (voir annexe 3 pour une illustration des changements réalisés sur les cartes). Dans les conditions « un seul changement par carte », les déplacements amènent l'étape changée à migrer d'une position pré- à une position post-changement qui était jusqu'alors inoccupée ; alors que dans la condition « six changements simultanés », les déplacements amènent les étapes changées à interchanger leur positions respectives (par exemple : après le changement, le préfecture occupe la position antérieurement occupée par l'église, qui occupe la position antérieurement occupée par la gare etc...).

Chaque carte correspond à un rang de changement spécifique (il y a 6 rangs de changements différents : n , $n + 1$, $n - 1$, $n > 1$, n_{final} , $global$, donc 6 cartes). L'ordre de passation des cartes est aléatoire.

⁸⁰ Cette caractéristique entraînera des analyses statistiques séparées pour chaque branche de départ.

L'affichage de signalisation des changements détectés est illustré en Figure 12 et consiste en la présentation de quatre propositions de réponse concernant le type de changement perçu (image en plus, en moins, déplacée, différente, ou « je ne sais pas ») et de toutes les icônes qui sont présentes dans le plan.

1.1.3. Équipement

L'expérience, pilotée par ordinateur, se déroule dans un box expérimental de l'Université de Provence. L'ordinateur utilisé pour la programmation, pour la passation de l'expérience, et pour l'enregistrement des données est un macintosh iBook, modèle 116498, équipé d'une souris. Les cartes de ville sont projetées, par une vidéo projecteur, sur un écran blanc de 100 * 80 cm, situé à 120 cm des participants. Les cartes ont été créées avec le logiciel Adobe Photoshop Elements®. L'expérience a été, quant à elle, programmée avec le logiciel Hypercard®.

1.1.4. Procédure : tâche de résolution de problème couplée à une tâche de détection des changements

L'expérience est passée individuellement, dans des conditions d'éclairage confortable. Les participants sont répartis en deux groupes, le groupe « déplacement » et le groupe « disparition », selon le type de changement auquel ils seront confrontés. Avant de commencer l'expérience, une consigne écrite informe les participants qu'ils vont incarner un contrebandier de cigares cubains, et que leur tâche sera de déplacer, le plus rapidement possible, un véhicule (contenant de la marchandise de contrebande) dans la carte d'une ville, afin d'acheminer leur marchandise jusqu'à destination (Paris). Pour cela, ils devront passer par un certain nombre d'étapes qui sont répertoriées dans un arbre des tâches, qui restera

affiché sur l'écran tout au long de l'expérience. Les participants sont également informés que des changements peuvent survenir sur la carte dans laquelle ils devront déplacer le véhicule, et qu'ils devront signaler tout changement aussitôt que perçu. Ils ne reçoivent aucune indication concernant le type de changement susceptible de survenir sur la carte.

Tout nouvel essai débute avec la présentation d'un écran de transition, et le click sur un bouton « Prêt » entraîne l'apparition d'une première carte sur l'écran. Au début de chaque essai, le véhicule à déplacer est situé à l'aéroport. La tâche des participants est de déplacer, avec la souris de l'ordinateur, ce véhicule dans le plan de la ville, conformément à un arbre des tâches, et en prenant soin de : (1) ne pas se rendre à une étape si les étapes qui lui sont sous-ordonnées n'ont pas été atteintes au préalable ; (2) ne pas sortir de la route sous peine d'être replacé automatiquement à la sortie de la dernière étape atteinte ; (3) atteindre l'étape but le plus rapidement possible. Quand le participant atteint une étape du trajet, un feed-back positif se fait entendre puis un masque visuel, d'une durée de 500 ms, recouvre brièvement la carte. Après la disparition du masque, la carte réapparaît sur l'écran et le véhicule est positionné à la sortie de l'étape atteinte. En revanche, quand le participant fait une sortie de route, ou qu'il atteint une étape incorrecte (c'est-à-dire une étape qui n'est pas répertoriée dans l'arbre des tâches -par exemple l'école -, ou alors une étape qui est répertoriée dans l'arbre des tâches mais dont l'atteinte à ce moment de la résolution de la tâche implique nécessairement qu'une des règles a été violée - par exemple, quand le participant tente d'atteindre une étape super-ordonnée alors qu'une des étapes qui lui est sous-ordonnée n'a pas encore été atteinte), un feed-back auditif négatif accompagne l'affichage d'un message d'erreur sur l'écran, et le véhicule est replacé à la sortie de la dernière étape atteinte correctement.

A un moment donné de la résolution du trajet, un changement est introduit sur l'une des étapes, pendant que la carte est cachée par le masque visuel. Ce changement (selon le

groupe : déplacement ou disparition de l'icône et du label d'une étape) reste sur l'écran jusqu'à ce que le participant le signale ou qu'il gagne l'étape but. Pour signaler un changement, le participant appuie sur la barre d'espace, ce qui entraîne l'apparition de l'écran de signalisation des changements, qui demande au participant, dans un premier temps, de désigner le type de changement qu'il a perçu (le participant doit choisir entre une image en plus, en moins, déplacée ou différente, et a, par ailleurs la possibilité de répondre qu'il ne sait pas quel le type de changement est survenu sur la carte), et dans un second temps, de déterminer l' (les) image(s) sur laquelle (lesquelles) le changement est survenu (le participant doit choisir, parmi les icônes de toutes les étapes qui figurent sur le plan celle(s) qu'il a perçue(s) comme ayant fait l'objet d'un changement). Puis, un bouton lui permet de revenir à la carte afin d'y terminer son trajet. Ainsi, après qu'un changement a été signalé (que ce signalement soit correct ou incorrect), l'étape changée réapparaît sur l'écran et/ou retrouve sa position initiale (selon que le changement a été une disparition ou un déplacement), et le participant retrouve son véhicule là où l'appui sur la barre d'espace en avait interrompu le déplacement. Enfin, quand le participant atteint l'étape but (Paris), un message indiquant que le trajet est terminé apparaît sur l'écran puis laisse place à l'écran de transition : l'appui sur un bouton « Prêt » entraîne l'apparition d'une nouvelle carte. Il en est de même pour toutes les cartes de l'expérience. L'expérience, qui s'achève quand les participants ont terminé de traiter les 6 cartes, dure approximativement 40 minutes. La Figure 12 donne un aperçu du déroulement de l'expérience.

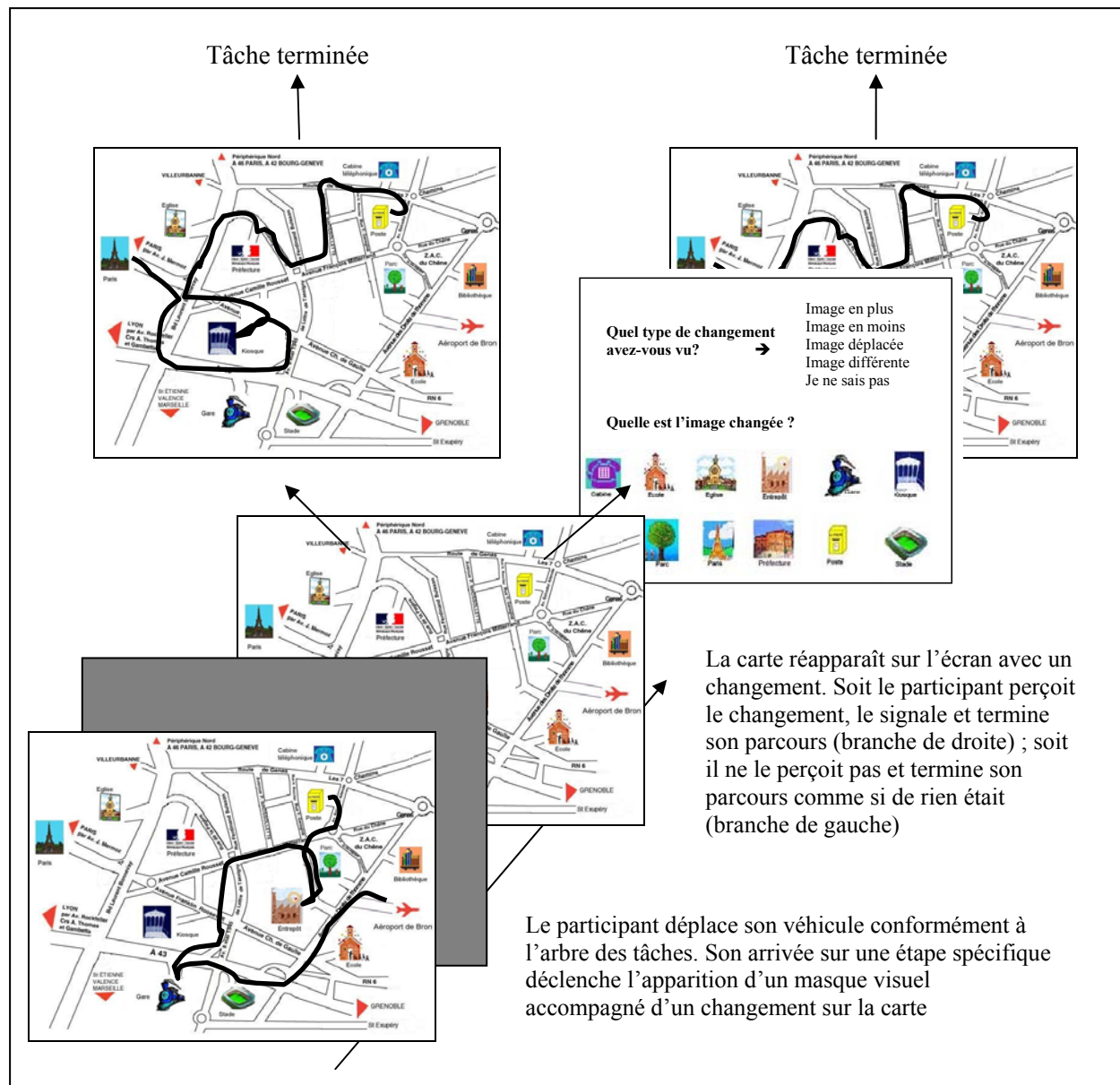


Figure 12. Déroulement des expériences 3, 4, et 5 (ici, le changement consiste en la disparition de l'entrepôt)

Préalablement à la passation de l'expérience, les participants passent une phase d'entraînement, qui est une version simplifiée de la tâche de l'expérience. La carte de l'entraînement met en scène quatre étapes, dont trois sont répertoriées dans l'arbre des tâches de l'entraînement. Dans un premier temps de la phase d'entraînement, l'expérimentateur explique comment l'arbre des tâches doit être lu et respecté, et ensuite fait une démonstration des différents parcours possibles, en testant toutes les possibilités de déplacements dans le

plan (déplacements licites et illicites, tels que les sorties de route, l'atteinte correcte et incorrecte d'une étape, l'atteinte du but, et la signalisation des changements), ainsi que leur conséquence sur le déplacement du véhicule. Ensuite, le participant exécute lui-même le parcours de l'entraînement deux fois sur cette même carte, en essayant de respecter la consigne de rapidité autant que faire se peut.

1.2. Analyse des résultats

Dans cette expérience, il s'agit d'étudier la détection des changements sur une scène visuelle complexe au fur et à mesure que la résolution de la tâche progresse. Aussi, les réponses des participants seront analysées en terme de nombre de détections correctes, en fonction de l'étape sous-but changée. Pour que leur réponse soit considérée comme correcte, les participants doivent avoir correctement identifié le type de changement et l'étape changée. Les autres réponses sont considérées comme incorrectes.

Étant donné que la situation expérimentale dans laquelle six changements (au lieu d'un seul) surviennent sur la carte diffère des cinq autres situations dans lesquelles un seul changement intervient, nous analyserons séparément les données recueillies dans les situations un et six changements.

1.2.1. Situations « un changement »

Plan d'analyse:

$\underline{S} < G_2 > * R_5$

Le facteur \underline{S} correspond aux participants ; le facteur G correspond au groupe (déplacement vs. disparition) ; le facteur R correspond au rang de l'étape sous-but affectée par le changement dans la résolution de la tâche ($n, n + 1, n - 1, n > 1, n_{final}$)

Avant de présenter nos résultats, il convient d'introduire la statistique particulière avec laquelle nous avons traité les données.

Les analyses statistiques traditionnellement utilisées en psychologie (ANOVAs, régressions linéaires...), dont certaines ont d'ailleurs été utilisées pour analyser les données des expériences 1 et 2, sont en fait des cas particuliers d'une classe de modèles appelés « modèles linéaires généraux ». Les modèles linéaires généraux permettent de tester les effets de variables explicatives catégorielles et continues dans des modèles contenant une ou plusieurs variables dépendantes qui doivent être *continues* et *distribuées normalement*. Dans sa forme la plus simple, un modèle général spécifie la relation linéaire qui existe entre une variable dépendante (Y) et un ensemble de variables explicatives (X_i), conduisant à l'équation linéaire :

$$Y = b_0 + b_1X_1 + b_2X_2 + \dots + b_kX_k + e$$

Dans cette équation, b_0 représente l'ordonnée à l'origine ; les valeurs b_i , les coefficients de régression des variables 1 à k calculés à partir des données ; et e , l'erreur de dispersion. Généralement dans les problèmes d'analyses des données, l'estimation des relations linéaires entre les variables permet de restituer de manière satisfaisante les données observées, et de prévoir raisonnablement de nouvelles observations. Toutefois, de nombreuses relations ne peuvent être synthétisées par une simple équation linéaire, ce qui est le cas lorsque la variable dépendante recueillie n'est pas distribuée normalement. Par exemple, si la variable dépendante est discrète, de type détecté (1) / pas détecté (0) comme c'est le cas dans les expériences 3, 4, et 5 de ce travail de thèse, alors les valeurs prévues par le modèle devraient

être également discrètes et toute autre valeur prévue serait illogique (cela n'aurait pas de sens que le modèle prédise une détection de 0.38, par exemple). Afin de traiter les variables dépendantes discrètes, il convient d'utiliser une extension des modèles linéaires généraux : les modèles linéaires généralisés. Les modèles linéaires généralisés sont utilisés pour prévoir les variations des variables dépendantes suivant des distributions discrètes (loi binomiale, multinomiale, poisson...) qui sont liées de façon non-linéaire aux variables explicatives, comme c'est le cas dans les expériences 3, 4, et 5. Ces modèles répondent à l'équation :

$$Y = g(b_0 + b_1X_1 + b_2X_2 + \dots + b_kX_k) + e$$

Dans cette équation, e représente l'erreur de dispersion et g la fonction de liaison, qui lie les effets des variables explicatives à la variable dépendante.

Les modèles linéaires généralisés diffèrent donc des modèles linéaires généraux par deux aspects majeurs : tout d'abord par la distribution de la variable dépendante, qui peut être *non-normale* et ne nécessite *pas* d'être continue, ensuite par une fonction de liaison, qui dépend de la distribution de la variable dépendante et qui décrit comment les valeurs prévues sont reliées aux variables explicatives linéaires. Classiquement, quand la distribution est binomiale, comme c'est le cas de notre variable dépendante, la fonction de liaison est une fonction logit, définie par l'équation :

$$\eta = \log(\mu / (1 - \mu))^{81}.$$

La procédure GENMOD⁸², que nous avons utilisée pour traiter nos données est aux modèles linéaires généralisés ce que l'ANOVA est aux modèles linéaires généraux. Cette

⁸¹ Notons que, dans les modèles linéaires généraux traditionnels, la fonction de liaison est une fonction identité, définie par l'équation : $\eta = \mu$

⁸² GENMOD pour « General Models », fonctionne sur le logiciel SAS®

procédure, qui teste la probabilité que $Y = 0$ par l'estimation d'un coefficient maximum de vraisemblance, permet de traiter les mesures répétées et d'analyser les contrastes.

Les résultats obtenus figurent dans le Tableau 3. Les valeurs indiquées dans le tableau sont exprimées en pourcentage pour permettre la comparaison entre les groupes d'effectifs différents, mais l'analyse statistique des données, à l'aide d'une procédure GENMOD, a été réalisée sur les données brutes (nombre de changements détectés).

Tableau 3 : Score de détection (%) des changements en fonction de la condition expérimentale et du rang de l'étape sous-but changée.

Rang de l'étape sous-but changée						
	n-1	n	n+1	n > 1	n final	<i>moyenne</i>
Condition expérimentale						
Déplacement	9.76	29.27	24.39	4.88	24.39	18.54
Disparition	26.92	80.77	69.23	80.77	57.69	63.08
<i>moyenne</i>	18.34	55.02	46.81	42.825	41.04	

Les analyses révèlent tout d'abord un effet simple du groupe sur le nombre de changements détectés ($X^2(1) = 17.20, p < .0001$) : les participants du groupe « disparition » détectent davantage de changements que les participants du groupe « déplacement » (le nombre de changements respectivement détectés est 82 / 130, soit 63.08% de détections, vs. 38 / 205, soit 18.54% de détections). Les analyses révèlent également un effet simple du facteur rang du changement ($n, n + 1, n - 1, n > 1, n_{final}$) sur le nombre de détection ($X^2(4) = 17.13, p = < .001$) : globalement, les changements sont mieux détectés quand ils surviennent sur l'étape impliquée dans le sous-but en cours de résolution (rang n) que quand ils surviennent sur une étape impliquée dans un autre sous-but. Enfin, les analyses révèlent un

effet de l'interaction du groupe et du rang du changement sur le nombre de détections de changements ($X^2(4) = 15.02, p < .005$), interaction qui se manifeste par un effet du rang du changement sur le nombre de détections plus marqué dans le groupe « déplacement » que dans le groupe « délétion ».

En ce qui concerne le groupe « déplacement », une analyse des contrastes révèle que les changements sont mieux détectés ($X^2(1) = 10.94, p < .0005$) quand ils surviennent sur les étapes impliquées dans le sous-but en cours de résolution (rangs n et $n + 1$) et le but final (rang n_{final}) que quand ils surviennent sur les autres étapes. Autrement dit, les changements de rang n et $n + 1$ sont mieux détectés que les changements de rang $n > 1$ (qui affectent une étape certes impliquée dans la résolution de la tâche, mais qui n'interviendra que bien après dans la procédure de résolution de la tâche), et que les changements de rang $n-1$ (qui affectent une étape qui était impliquée dans un sous sous-but antérieurement résolu). Plus précisément, les analyses révèlent que les changements de rang n , qui affectent l'étape immédiatement impliquée dans le sous-but en cours de résolution, sont mieux détectés que les changements qui surviennent sur les étapes impliquées dans un autre sous-but ($X^2(1) = 3.89, p < .05$).

En ce qui concerne le groupe « disparition », les analyses montrent que les changements de rang $n, n + 1, n > 1$, et n_{final} (autrement dit les changements qui affectent une étape sous-but impliquée dans l'atteinte du but et par lesquelles les participants sont obligés de passer) sont mieux détectés que les changements de rang $n-1$ (qui affectent une étape qui était impliquée dans un sous-but antérieurement résolu, et donc par laquelle les participants n'auront plus à passer pour parvenir à l'étape but) ($X^2(1) = 10.59, p = .001$). Rappelons, en effet, que l'expérience a été programmée de telle sorte que les changements restent sur l'écran jusqu'à ce que les participants les signalent ou qu'ils aient terminé la tâche, c'est-à-dire atteint l'étape but (Paris). Ainsi, dans le groupe « disparition », quand les étapes sous-but sont effacées du plan, elle demeurent absentes du plan tant que le participant n'a signalé aucun

changement et n'a pas non plus atteint l'étape but. Or pour atteindre l'étape but, les participants sont obligés de passer par ces mêmes étapes qui ne figurent plus sur la carte (*ie.*, les étapes impliquées dans les sous-buts de rang : $n, n+1, n > 1, n_{final}$). En quelque sorte, dans le groupe « disparition », les participants sont quasi-obligés de détecter les changements, puisque les étapes qu'ils doivent atteindre ne figurent plus sur le plan après que le changement ait eu lieu. Dans cette condition, il est alors intéressant d'étudier non plus combien de changements sont détectés en fonction du rang du changement mais plutôt à quel moment, dans la procédure de résolution du trajet, les participants perçoivent et signalent les disparitions d'étapes. Soit la détection de ces changements, particulièrement saillants du point de vue perceptif⁸³, dépend uniquement de leur saillance perceptive, auquel cas l'on devrait observer que les détections des disparitions d'étapes ne sont pas liées à la résolution de la tâche (modèle perceptif) ; soit, au contraire, les participants ne perçoivent les disparitions d'étapes uniquement lorsque les étapes effacées du plan s'avèrent nécessaires pour résoudre le sous-but dans lequel ils sont engagés, et cela indépendamment de la saillance perceptive de ces changements (modèle fonctionnel). Dans ce cas, l'on devrait observer que les changements sont signalés au fur et à mesure que les étapes disparues deviennent nécessaires au sous-but en cours de résolution. Par conséquent, si la détection des disparitions d'étapes suit la procédure de résolution de la tâche, un modèle fonctionnel prédit que les changements sont détectés au fur et à mesure qu'ils deviennent nécessaires à la résolution du sous-but dans lequel les participants sont engagés (*ie.*, quand le véhicule des participants atteint l'étape qui précède l'étape disparue, dans la résolution du trajet). Rappelons que les différents changements sont déclenchés à un temps t de la résolution de la tâche par l'arrivée du véhicule du participant sur des étapes spécifiques (cf. Figure 11). On devrait donc observer

⁸³ Les disparitions d'étapes laissent en effet un « trou » dans la carte, là où jadis figurait une étape

que les changements de rang n (qui altèrent un étape qui doit être immédiatement atteinte) seront détectés au temps t (*ie.*, dès que introduits dans la carte) ; que les changements de rang $n + 1$ seront détectés au temps $t+1$; que les changements de rang $n > 1$ seront détectés au temps $t+3$; et que les changements de rang n_{final} seront détectés au temps $t+3$. La Figure 13 illustre ces prédictions concernant le moment de la détection du changement dans le groupe disparition.

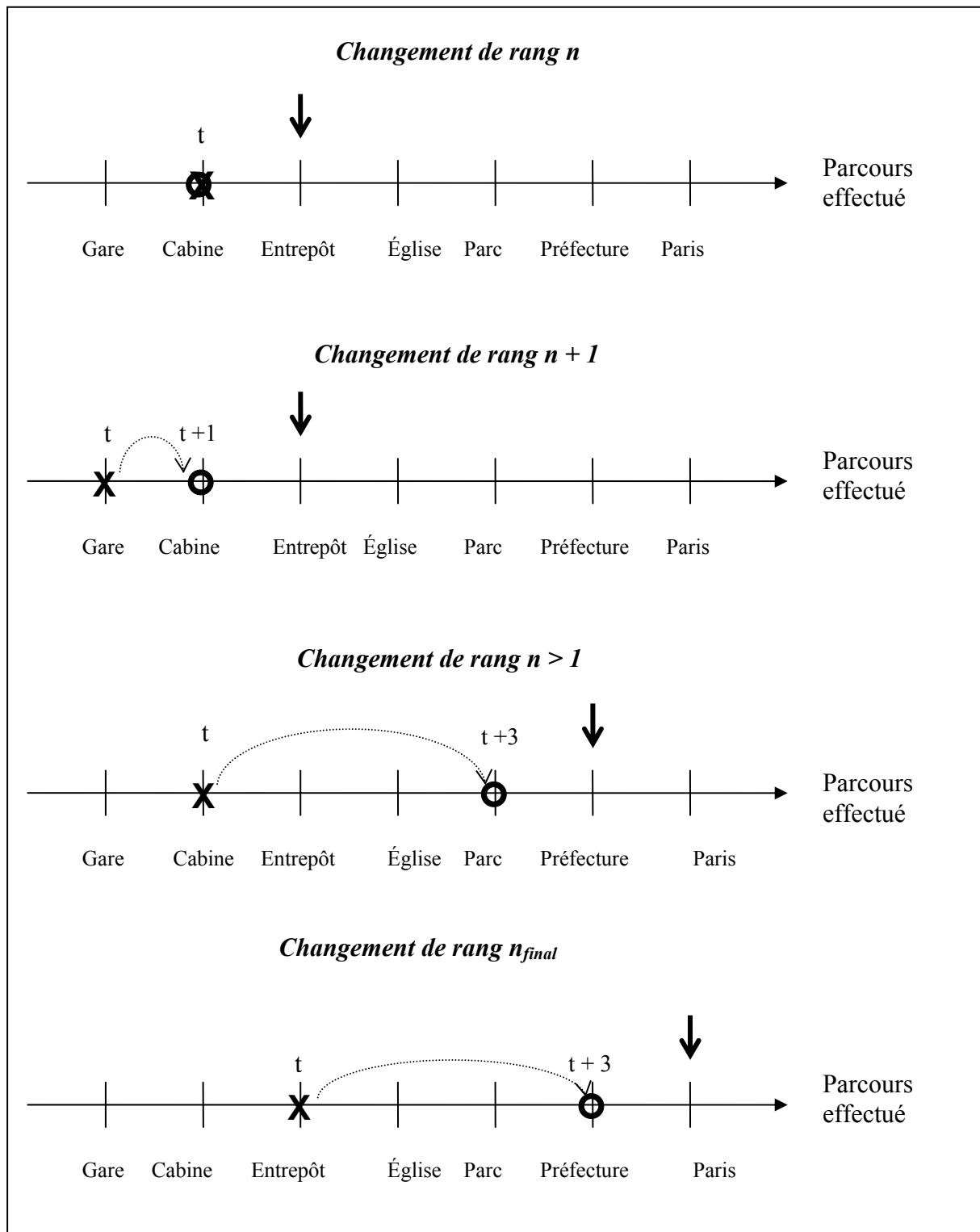


Figure 13. Prédications sur le moment de la détection des changements, dans la condition « disparition », selon le modèle fonctionnel qui stipule que la détection des changements dépend de la tâche. La croix représente l'étape atteinte par le participant, la flèche (en caractère gras) désigne l'étape modifiée et t représente le moment où le changement apparaît sur la carte. Le cercle donne le moment de détection des changements prédit par le modèle, pour un participant qui effectuerait le trajet : gare → cabine → entrepôt → église → parc → préfecture → paris.

Afin de tester ces prédictions, nous avons catégorisé les détections correctes des participants du groupe « disparition », selon que ces détections surviennent quand l'étape disparue devient nécessaire à la résolution du sous-but en cours (tel que prédit par le modèle en Figure 13) ou non. Les résultats obtenus figurent dans le Tableau 4⁸⁴.

Tableau 4. Effectifs observés dans le groupe « disparition », selon que le moment de la détection des changements est conforme ou non aux prédictions du modèle fonctionnel, en fonction du rang des changements.

	Moment de la détection du changement	
	Conforme	Non conforme
Rang du changement		
$n + 1$	16	2
$n > 1$	12	3
n_{final}	14	7

Un test de Khi-deux révèle que les participants tendent à signaler le changement au moment où l'étape changée devient pertinente par rapport au sous-but en cours de résolution ($X^2(2) = 132.00, p < .00001$). Les résultats obtenus sont donc conformes aux prédictions du modèle fonctionnel, c'est-à-dire que, dans la condition « disparition », les changements sont perçus au fur et à mesure que la résolution de la tâche progresse et que les étapes sous-but altérées par le changement s'avèrent pertinentes par rapport au sous-but en cours de résolution.

⁸⁴ Nous n'avons pas fait figurer la condition « changement de rang n » dans ce tableau, car dans cette condition, les participants n'ont pas d'autre alternative que de détecter le changement au moment où celui-ci est nécessaire à la résolution de la tâche, conformément aux prédictions du modèle. En effet, dès sa survenue, le changement affecte l'étape de rang n : les participants sont obligés de le signaler aussitôt, sinon, ils se voient dans l'incapacité de continuer leur trajet et d'atteindre l'étape but.

1.2.2. Situation « six changements simultanés »

Plan d'analyse :

$$\underline{S} < G_2 * B_2 >$$

Le facteur \underline{S} correspond aux participants ; le facteur G correspond au groupe (déplacement vs. disparition) ; le facteur B correspond à la branche de départ (« entrepôt » vs. « préfecture »)

Dans cette condition, nous avons voulu étudier la détection de changements multiples introduits dans les cartes. Le fait que les changements sont déclenchés par l'arrivée du véhicule du participant sur l'entrepôt, quelle qu'ait été la branche de départ (« entrepôt » ou « préfecture »), et à l'intérieur de chaque branche, quelle qu'ait été l'étape de départ (« gare » ou « cabine » pour la branche « entrepôt » ; « église » ou « parc » pour la branche « préfecture »), nécessite que soient conduites des analyses séparées pour chaque branche. Dans la condition « 6 changements simultanés », la branche « entrepôt » est la branche de départ pour 18 des 26 participants du groupe « disparition » et pour 34 des 41 participants du groupe « déplacement » ; les autres participants commençant leur trajet par la branche « préfecture ». Ce biais dans le choix de la branche de départ a pour conséquence un nombre insuffisant de données pour les participants qui commencent par la branche « préfecture », aussi nous ne conduirons d'analyses (test de khi-deux) qu'à partir des données des participants qui commencent leur parcours par la branche « entrepôt ».

Les résultats montrent dans un premier temps que les participants détectent, en moyenne 1.68 changements (avec, au maximum, quatre changements détectés sur les six introduits dans les scènes) quand ils appartiennent au groupe « disparition » et seulement 0.63 changements (avec un maximum de deux changements détectés sur les six introduits dans les

scènes) quand ils appartiennent au groupe « déplacement »⁸⁵. De plus, les résultats montrent que 72% des participants du groupe « disparition » détectent *moins* de quatre changements sur six, et que 94% des participants du groupe « déplacement » détectent *moins* de deux changements sur six.

Il est intéressant, dans cette condition, d'observer quelles sont les étapes dont les changements ont été détectés. Dans une perspective fonctionnelle, où la détection des changements serait liée à la tâche, nous devrions observer que les changements sont d'autant mieux détectés qu'ils concernent les étapes immédiatement nécessaires à la résolution du sous-but dans lequel les participants sont engagés.

Représentons linéairement le parcours des participants et nommons « x » l'étape les participants doivent atteindre : nous avons alors « $x + 1$ », qui représente l'étape qui deviendra nécessaire quand les participants auront atteint l'étape x , « $x - 1$ » qui représente l'étape qui devient inutile quand les participants ont atteint l'étape x , et ainsi de suite pour « $x + 2$ », « $x - 2$ »... En suivant ce raisonnement, nous avons recodé toutes les étapes à partir de l'étude des protocoles individuels. Une perspective fonctionnelle prédit que, dans cette condition « 6 changements simultanés », les changements devraient être détectés d'autant plus facilement qu'ils altèrent l'étape x , alors qu'une perspective purement perceptive, qui suppose que la détection des changements n'est pas fonction de la résolution de la tâche, prédit que tous les changements ont autant de chances d'être détectés. Nous avons donc testé⁸⁶, pour les deux groupes, dans quelle mesure la distribution de la détection des changements, en fonction de l'ordre de l'étape changée dans la résolution de la tâche, est en adéquation avec ce que

⁸⁵ De plus, si l'on considère toutes les réponses recueillies dans la situation « six changements simultanés » (correctes et incorrectes), le nombre moyen d'étapes *signalées* comme ayant subi un changement est de 2.4 pour le groupe « disparition » et 1.3 pour le groupe « déplacement », ce qui reste très inférieur aux six changements effectivement réalisés sur la scène visuelle (voir Hayhoe et al., 1998 pour des résultats similaires).

⁸⁶ Par un test de khi-deux

prédirait le hasard (ou selon une perspective purement perceptive, dans laquelle les détections se distribueraient de manière homogène quelle que soit l'ordre de l'étape changée, *ie.*, x ; $x + 1$; $x - 1$...). Les analyses révèlent que la distribution des détections des changements diffère significativement de ce que prédirait le hasard, autant dans le groupe « disparition » ($X^2(5) = 17.50$, $p < .005$) que dans le groupe « déplacement » ($X^2(5) = 34.60$, $p < .00005$), comme le montre la Figure 14.

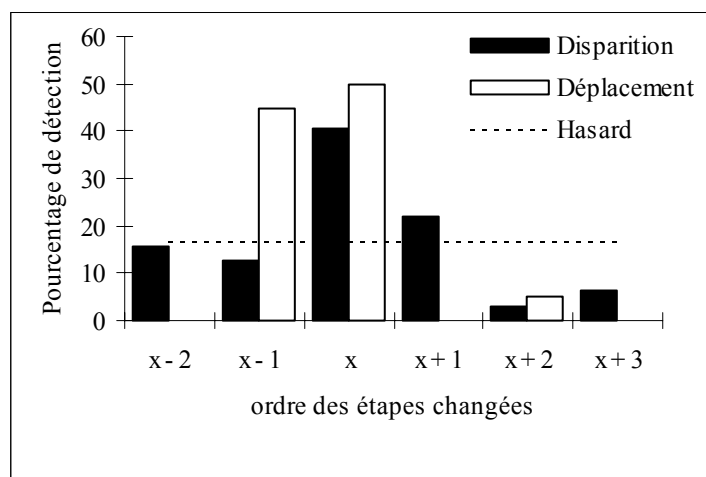


Figure 14. Répartition des détections des changements (%) en fonction du groupe et de l'ordre de l'étape sous-but changée dans la structure des sous-buts. La ligne discontinue représente la répartition des changements prédite par le hasard.

Dans les deux groupes, les changements qui surviennent sur les étapes immédiatement nécessaires à la résolution du sous-but dans lequel les participants sont engagés (étapes x) sont mieux détectés que ce que prédit le hasard, alors que ce n'est pas le cas des autres changements (exception faite des changements survenant sur les étapes $x-1$, dans le groupe « déplacement »⁸⁷).

⁸⁷ Rappelons que dans la situation « six changements simultanés », les positions des six étapes sont interchangeables. Or, une analyse des protocoles individuels révèle que, pour plus de la moitié des participants du groupe « déplacement » (20 participants sur 34), le changement mutuel des positions de deux étapes s'avère concerner, *in fine*, les étapes x et $x+1$. Dans ce cas, il est tout à fait possible que les participants aient dirigé leur véhicule là où aurait dû figurer l'étape x , et qu'ils aient trouvé à la place l'étape $x-1$. Par la suite, l'on imagine aisément qu'ils aient signalé le changement des étapes x et $x-1$.

1.3. Discussion

Dans cette expérience, le but de la tâche assignée aux participants est d'atteindre, le plus rapidement possible, une étape but (Paris). Or, pour atteindre cette étape but, les participants doivent préalablement satisfaire deux sous-buts (atteindre l'entrepôt et la préfecture). La réalisation de ces sous-buts nécessite, de la même manière, que soient atteintes deux autres étapes (la gare et la cabine pour le sous-but entrepôt ; l'église et le parc pour le sous-but préfecture). Durant ce parcours, un changement survient sur l'une des étapes pertinentes par rapport à la tâche, et l'on utilise la détection des changements comme un indicateur des informations de la scène qui ont été focalisées par l'attention et représentées en mémoire.

Les résultats montrent que la tâche influence la représentation des informations visuelles en mémoire : le contenu des représentations visuelles évolue au fur et à mesure que la résolution des sous-buts progresse. En effet, l'analyse des données recueillies dans les situations « un changement » montre que les changements sont mieux perçus par les participants du groupe « déplacement » quand ils sont directement pertinents par rapport à la tâche. En effet, les participants détectent davantage un changement qui survient sur une étape impliquée dans le sous-but en cours de résolution (ie. une étape de rang n ou $n + 1$) et dans le but final qu'ils ne détectent un changement qui survient sur des étapes qui ne sont plus pertinentes par rapport à la tâche (rang $n - 1$) ou qui, bien que toujours pertinentes par rapport à la tâche, appartiennent à un sous-but ultérieur trop éloigné du sous-but en cours de résolution dans la structure de sous-buts (rang $n > 1$). Ainsi, il semblerait que les informations visuelles sont d'autant mieux représentées en mémoire qu'elles sont pertinentes par rapport à la tâche et immédiatement impliquées dans le sous-but en cours de résolution. Ce résultat se retrouve, sous une forme un peu différente, dans le groupe « disparition ». En effet, dans ce

groupe, les résultats révèlent un effet du rang du changement sur leur détection qui contraste avec l'effet observé dans le groupe « déplacement » : les changements les mieux perçus sont ceux qui affectent les étapes impliquées dans la résolution de la tâche (*ie.*, les changements de rang n , $n+1$, $n > 1$, et n_{final}), par opposition aux changements de rang $n - 1$ qui concernent une étape sous-but déjà atteinte et donc dorénavant inutile à la résolution de la tâche. Ce résultat s'explique par une caractéristique propre à l'expérience : les changements restent visibles sur les cartes jusqu'à ce qu'ils soient signalés ou que les participants terminent la tâche. Dans le cas des disparitions d'étapes, les étapes disparues restent donc absentes de la carte tant que le participant n'en a pas signalé le changement ou atteint l'étape but. Or, les participants ne peuvent pas atteindre l'étape but s'ils n'ont pas au préalable atteint les étapes absentes de la carte! Ainsi, la condition « disparition » entraîne de façon quasi-obligatoire la détection des étapes effacées, ce qui est confirmé par nos résultats. Nous avons donc étudié, dans cette condition « disparition », à quel moment les changements étaient perçus par les participants, afin de déterminer si la détection des changements restait tout de même liée à la résolution de la tâche. Les résultats montrent que les changements sont perçus à partir du moment où le sous-but qui met en cause l'étape changée devient le sous-but en cours de résolution. Ainsi, les représentations visuelles fonctionnelles sont influencées par la résolution de la tâche et contiennent préférentiellement les informations visuelles directement et immédiatement nécessaires à la résolution du sous-but dans lequel nous sommes engagés. Ces représentations sont donc transitoires et dynamiques puisque les informations qu'elles contiennent sont sans cesse actualisées au fur et à mesure que la résolution de la tâche progresse.

Dans une certaine mesure, il peut sembler surprenant que le but de la tâche soit, dans le groupe « déplacement », presque aussi bien représenté que les étapes impliquées dans le sous-but en cours de résolution, ce qui tendrait à suggérer que les représentations visuelles contiendraient en fait des informations transitoires (les informations pertinentes par rapport au

sous-but en cours, et qui sont donc sans cesse renouvelées au fur et à mesure de la résolution des sous-buts) et des informations plus durables, relatives au but ultime de la tâche. Il est donc possible que le but de la tâche ait été codé différemment des sous-buts, ne serait-ce que parce qu'il permet de construire la structure des buts, c'est-à-dire la suite des sous-buts engendrée à partir du but principal (Richard, 1999), et définit la configuration finale de la tâche. Par conséquent, le but de la tâche jouirait d'un traitement attentionnel particulier *avant même* de commencer la résolution de la tâche, et les résultats observés dans le groupe « déplacement » concernant la représentation visuelle du but de la tâche signeraient alors l'existence de post-effets attentionnels sur les représentations en mémoire des informations liées au but de la tâche.

Cette interprétation est néanmoins à considérer avec prudence : dans les cartes utilisées dans cette expérience figurait des indications de directions, et la localisation de « Paris » (l'étape but) était indiquée par une flèche (cf. Figure 9, plus haut dans le texte). Il est donc tout à fait possible que les bonnes détections du déplacement de l'étape représentant Paris aient été facilitées par ces indications de direction puisque la position post-changement de l'étape but « Paris » est en désaccord avec la flèche indiquant la direction de Paris. Imaginons un instant que les participants se servent de l'information fléchée pour se diriger vers l'étape but : ils se rendent nécessairement compte que la direction indiquée ne conduit pas à Paris, et signalent le déplacement de l'étape. La détection de ce changement n'a alors plus rien à voir avec l'influence du but de la tâche sur la représentation des informations visuelles en mémoire. Cette explication peut également tout à fait rendre compte, chez le groupe « disparition », du moindre pourcentage de détection des changements de rang n_{final} (57.69%) par rapport aux changements de rang $n > 1$ (80.77%), alors que les deux changements surviennent sur des étapes impliquées dans la réalisation de la tâche. En effet, en dépit de sa disparition, la localisation de l'étape but est indiquée sur la carte et les participants peuvent

néanmoins s'y rendre et terminer la tâche, sans même s'être aperçus que l'icône de l'étape n'était plus présente sur la carte⁸⁸.

Il n'est, pour l'instant, malheureusement pas possible de départager ces deux explications.

L'analyse de la situation « six changements simultanés » révèle quant à elle un fait intéressant : les participants ont tendance à ne détecter que deux ou trois changements sur les six qui altèrent effectivement la carte, ce qui est plutôt surprenant compte tenu du fait que la modification globale de la carte qui en résulte est particulièrement saillante du point de vue perceptif (surtout pour le groupe disparition ; voir annexe 4, pour une illustration de la disparition de six étapes d'une carte). Quand on s'intéresse au détail de ces détections, on remarque qu'elles concernent majoritairement les étapes qui sont immédiatement impliquées dans la résolution du sous-but dans lequel les participants sont engagés. En effet, les participants signalent davantage le changement des étapes directement pertinentes par rapport au sous-but en cours de résolution, et de façon moindre le changement des autres étapes, que ne le prédirait le hasard. Ce résultat semble suggérer que même lorsqu'on renforce la saillance perceptive des changements qui altèrent la scène visuelle, seules les informations visuelles directement requises par la tâche sont actualisées et comparées aux entrées sensorielles. Ce mode de traitement progressif des informations visuelles de la scène serait robuste et insensible à l'augmentation de la stimulation sensorielle, puisque, même considérées ensemble, les réponses (correctes et incorrectes) des participants aux six changements simultanés sous-estiment toujours l'ampleur réelle de l'altération survenue sur la scène visuelle, et révèlent ainsi cruellement les limitations de la mémoire de travail visuelle.

⁸⁸ Le programme de l'expérience réagissant à l'arrivée du véhicule sur certaines coordonnées, il ne traite malheureusement pas le fait que l'icône de l'étape atteinte soit présente ou absente de la carte.

En conclusion, la perception d'une scène visuelle complexe en condition de vision active, c'est-à-dire lorsque les participants sont engagés dans l'accomplissement d'une tâche sur une scène visuelle, consisterait en l'extraction, la représentation et l'actualisation progressive de l'information visuelle au fur et à mesure que celle-ci devient nécessaire à la réalisation du sous-but en cours de résolution. Les représentations visuelles fonctionnelles de la scène seraient donc déterminées par les opérations cognitives entreprises, et ne serviraient pas de façon globale la tâche prise dans son ensemble, indépendamment des besoins immédiats de l'observateur. Une fois qu'une opération cognitive serait terminée, la représentation visuelle des informations visuelles associées déclinerait progressivement en mémoire, et serait remplacée par la représentation de nouvelles informations, associées à la nouvelle opération cognitive entreprise. Les représentations visuelles de la scène seraient donc hautement fonctionnelles et transitoires, en ce sens où, bien que constamment accessibles, les informations visuelles portées par la scène ne seraient réellement traitées par le système cognitif que lorsqu'elles seraient immédiatement requises par les demandes de la tâche : la représentation visuelle de la scène serait redessinée à chaque fois qu'un nouveau sous-but prendrait le pas sur un ancien, conduisant à autant de représentations *différentes* d'une *même* scène, qu'il y aurait de sous-buts à résoudre. L'actualisation et la comparaison des informations pré- et post-changements répondrait à un critère d'immédiateté des besoins fonctionnels, et ne serait pas facilitée par la circulation prolongée (au propre comme au figuré en ce qui concerne notre expérience) dans la scène, ni par le renforcement de la saillance perceptive des changements.

Néanmoins, il reste possible qu'au-delà du traitement privilégié des informations visuelles immédiatement nécessaires et requises par la tâche, les résultats obtenus ne résultent que du simple fait que les informations visuelles n'aient pas été fovéalisées avant et après le changement, ce qui rendrait impossible la détection des changements, comme le supposent

Hollingworth et collaborateurs (Hollingworth & Henderson, 2002; Hollingworth et al., 2001). Ces derniers considèrent en effet que les représentations visuelles d'une scène seraient relativement détaillées, et que la cécité au changement observée dans les paradigmes avec report explicite du changement résulterait, en fait, de la non fovéalisation des objets *avant* et *après* le changement, invalidant le processus de comparaison nécessaire à la détection des changements. Bien que notre position considère que l'effet de la tâche joue dès la prise d'information sur la scène, elle suppose également que les représentations visuelles sont construites par et pour la réalisation de la tâche, et que la cécité au changements des étapes qui ne sont pas immédiatement nécessaires à la résolution de la tâche serait répliquée même si l'on forçait la fovéalisation de tous les objets de la scène *avant* le changement. Plusieurs de nos résultats suggèrent déjà que l'exploration prolongée de la scène (physique du point de vue du déplacement du véhicule au sein de la carte ; et cognitive du point de vue de l'évolution dans la structure des sous-buts) ne favorise pas l'intégration des informations visuelles en mémoire. En effet, si l'on prend l'exemple des changements de rang $n > 1$, la représentation de l'étape ne s'enrichit pas du temps passé à circuler dans la scène : rares sont les changements de rang $n > 1$ à être détectés en condition « déplacement ». Pourtant, il y a de grandes chances pour que, au cours de leur parcours, les participants soient passés plusieurs fois à proximité de l'étape sous-but de rang $n > 1$ sans que ces visites inopinées aient entraîné l'accumulation en mémoire des informations visuelles relatives à cette étape, ce qui aurait sinon entraîné la réussite du processus de comparaison pré- et post-changement à point nommé et la détection du changement de l'étape. Néanmoins, afin de déterminer formellement si la cécité au changement observée dans la présente expérience résulte de représentations transitoires de la scène comme nous le supposons, ou d'un défaut de fovéalisation de l'élément changé en dépit de représentations détaillées comme le supposent Hollingworth et collaborateurs (Hollingworth & Henderson, 2002; Hollingworth et al., 2001),

nous avons, dans les expériences suivantes, forcé la fovéalisation de toutes les étapes présentes dans la scène *avant* que les participants commencent la tâche. Si les résultats observés dans l'expérience 3 résultent du simple fait que seuls les éléments pertinents par rapport à la tâche sont fovéalisés et que la détection de leur changement s'en retrouve facilitée, alors l'on ne devrait plus observer d'effet de la tâche sur la détection des changements. Au contraire, si cet effet est répliqué, alors l'on pourra penser que la cécité au changement résulte de représentations visuelles fonctionnelles et transitoires.

2. EXPÉRIENCE 4 : EFFET DU RANG DU SOUS-BUT MODIFIÉ SUR LA DÉTECTION DES CHANGEMENTS : UN EFFET DE L'EXPLORATION VISUELLE (PARTIE 1) ?

Dans cette expérience, il s'agit de contraindre les participants à concentrer leur regard sur toutes les étapes contenues dans la carte avant de commencer leur parcours, afin de vérifier si l'effet de la tâche sur la détection des changements observé dans l'expérience 3 est répliqué alors que toutes les étapes ont été perçues *avant* le changement. Répliquer cet effet suggèrerait que la cécité au changement observée dans l'expérience 3 résulte bel et bien de représentations visuelles fonctionnelles transitoires, ne codant que les informations visuelles immédiatement nécessaires à la résolution de la tâche, le temps qu'elles sont nécessaires. Le cas inverse nous conduirait plutôt à envisager que les représentations des scènes visuelles complexes sont relativement détaillées dès lors que les informations visuelles qui les composent ont été fovéalisées : ces représentations ne seraient donc pas forcément liées à la tâche.

2.1. Méthode

2.1.1. *Participants*

Vingt-quatre étudiants inscrits en licence de Psychologie de l'Université de Provence ont pris part à cette expérience. Tous ont attesté d'une acuité visuelle normale ou corrigée et étaient naïfs du point de vue des objectifs de cette recherche. Aucun d'entre eux n'a participé aux expériences 1, 2 ou 3.

2.1.2. *Matériel*

Le matériel (les cartes de la ville, l'écran de signalisation des changements ainsi que l'arbre des tâches) utilisé dans l'expérience et l'entraînement est le même que celui de l'expérience 3, à ce détail près que nous avons effacé des cartes toutes les indications de directions, dont nous pensions qu'elles pouvaient avoir été à l'origine de stratégies de détection dans l'expérience 3. Dans l'expérience 4, tous les changements sont des déplacements, et nous avons remplacé la situation « six changements simultanés » par une situation dans laquelle le changement est non pertinent (NP), c'est-à-dire qu'il survient sur une étape qui n'est pas référencée dans l'arbre des tâches.

L'ordre de passation des cartes est aléatoire.

2.1.3. *Équipement*

L'équipement utilisé est le même que celui de l'expérience 3.

2.1.4. Procédure : tâche de résolution de problème couplée à une tâche de détection des changements

La procédure utilisée est, à peu de choses près, la même que celle de l'expérience 3. Nous ne présenterons donc ici que les modifications apportées par l'expérience 4. Afin de mesurer la détection de changement en fonction du rang de l'étape modifiée dans la structure des sous-buts, nous n'avons utilisé qu'un seul type de changement : les déplacements d'étapes⁸⁹. Nous avons aussi remplacé la situation « six changements simultanés » par une situation expérimentale dans laquelle un seul changement survient sur la carte : ce changement est un changement non pertinent, c'est-à-dire qu'il altère une étape de la carte qui ne figure pas dans l'arbre des tâches (le kiosque, cf. annexe 3).

Comme dans l'expérience 3, tout nouvel essai débute avec la présentation d'un écran de transition, et l'appui sur un bouton « Prêt » entraîne l'apparition d'une première carte sur l'écran, carte de laquelle le véhicule des participants est absent, et sur laquelle toutes les étapes (qu'elles soient ou non répertoriées dans l'arbre des tâches) sont masquées par un cache noir. La nouveauté de l'expérience 4, est que, avant de commencer à déplacer le véhicule dans la carte, les participants doivent démasquer toutes les étapes en cliquant sur chacun des caches, ce qui nous assure que les icônes de toutes les étapes ont été perçues avant que les participants ne commencent la tâche. Une fois que toutes les étapes ont été dévoilées, le véhicule apparaît sur l'aéroport, et les participants entament leur parcours. Ensuite, l'expérience se déroule comme dans l'expérience 3.

⁸⁹ Ce choix a été motivé par le fait que les déplacements d'étapes nous semblaient constituer un bon moyen d'évaluer les représentations visuelles, par rapport aux disparitions qui sont, comme nous l'avons vu, des changements à détection « quasi obligée » pour que les participants puissent terminer la tâche.

2.2. Analyse des résultats

Plan d'analyse:

$\underline{S}_{24} * R_6$

Le facteur \underline{S} correspond aux participants ; et le facteur R correspond au rang de l'étape sous-but affectée par le changement dans la résolution de la tâche ($n, n + 1, n - 1, n > 1, n_{final}, NP$)

Les données ont été analysées au moyen d'une procédure GENMOD, et les résultats figurent dans le Tableau 5.

Tableau 5. Score de détection (%) des changements en fonction du rang de l'étape sous-but changée

Rang de l'étape changée					
$n-1$	n	$n+1$	$n > 1$	n_{final}	NP
33.33	45.83	37.50	16.67	33.33	12.50

Les résultats ne révèlent pas d'effet de la tâche sur la détection des changements ($X^2 (5) = 7.66, ns.$).

2.3. Discussion

L'expérience 4 ne réplique pas les résultats de l'expérience 3, ce qui tend à suggérer qu'une fois que l'on a forcé la focalisation oculaire sur les icônes de toutes les étapes comprises dans le plan, les participants sont en mesure de comparer l'information visuelle pré- et post-changement, et détectent les changements indépendamment de leur pertinence par rapport à la tâche. On remarque toutefois que l'on retrouve les meilleures performances de

détection pour les changements de rang n , qui altèrent l'étape sous-but en cours de résolution, et que les performances sont d'autant plus mauvaises que les changements altèrent des étapes qui s'en éloignent. Aussi, peut-être faudrait-il considérer cette absence d'effet significatif de la tâche sur la détection des changements non pas comme la signature de représentations visuelles affranchies de l'influence de la tâche, mais plutôt comme le résultat de l'effet interférent d'une double tâche, que les participants auraient eue à réaliser sur les scènes. En effet, la tâche de démasquage peut avoir été interprétée comme une première tâche à réaliser sur les scènes, d'autant que la consigne stipule que le démasquage des étapes rendra possible le déplacement du véhicule. Cette première tâche aurait attribué un statut de pertinence à *toutes* les étapes à démasquer qui auraient été par conséquent traitées et représentées en mémoire, ce qui aurait pu interférer avec le statut de pertinence attribué par la (vraie) tâche aux étapes de la scène répertoriées dans l'arbre des tâches, conduisant ainsi à annihiler l'effet observé dans l'expérience 3. C'est pourquoi, dans le but de déterminer définitivement si l'effet observé dans l'expérience 3 est ou non le fait de représentations visuelles fonctionnelles et transitoires, nous avons repris les principes et objectifs de cette expérience dans une cinquième expérience en prenant soin de ne pas introduire de double tâche sur les scènes.

3. EXPÉRIENCE 5 : EFFET DU RANG DU SOUS-BUT MODIFIÉ SUR LA DÉTECTION DES CHANGEMENTS : UN EFFET DE L'EXPLORATION VISUELLE (PARTIE 2) ?

Afin de nous prémunir de l'effet potentiel de la double tâche, nous avons repris le principe de la focalisation des participants sur toutes les étapes de la carte préalablement à l'exécution du trajet, mais en considérant les étapes non plus comme des parties de la carte à démasquer, mais simplement comme un fond dans lequel est dissimulée une forme (un point rouge) que les participants doivent localiser le plus rapidement possible.

3.1. Méthode

3.1.1. Participants

Vingt-quatre étudiants inscrits en licence de Psychologie de l'Université de Provence ont pris part à cette expérience. Tous ont attesté d'une acuité visuelle normale ou corrigée et étaient naïfs du point de vue des objectifs de cette recherche. Aucun d'entre eux n'a participé aux expériences 1, 2, 3 ou 4.

3.1.2. Matériel

Le matériel (les cartes de la ville, l'écran de signalisation des changement ainsi que l'arbre des tâches) utilisé dans l'expérience et l'entraînement est le même que celui de

l'expérience 4. Comme dans cette dernière, tous les changements sont des déplacements, et l'on manipule six rangs de changements : $(n ; n + 1 ; n - 1 ; n > 1 ; n_{final} ; \text{et } NP)$.

Le point rouge est toujours caché à l'intérieur de l'icône d'une étape de la carte. Les étapes qui contiennent le point sont déterminées pseudo aléatoirement : une même étape d'une même carte ne peut pas contenir plus d'une fois le point rouge.

L'ordre de passation des cartes est aléatoire.

3.1.3. *Équipement*

Le matériel utilisé est le même que celui de l'expérience 3 et 4.

3.1.4. *Procédure : tâche de résolution de problème couplée à une tâche de détection des changements*

Comme dans l'expérience 4, tout nouvel essai débute avec la présentation d'un écran de transition, et l'appui sur un bouton « Prêt » entraîne l'apparition d'une première carte sur l'écran, carte de laquelle le véhicule des participants est absent. Les participants doivent trouver le plus rapidement possible un point rouge caché dans la carte. Quand ils l'ont trouvé, ils doivent cliquer sur l'icône qui contient le point : la carte disparaît brièvement et réapparaît, contenant un nouveau point à localiser, et ainsi de suite jusqu'à ce que les participants aient déniché quatre points. Étant donné que les points ne peuvent être localisés deux fois dans une même étape pour une même carte, l'on peut penser que la recherche visuelle des quatre points sur une même carte conduit les participants à balayer l'ensemble de la scène et à concentrer au moins une fois leur regard sur chacune des étapes de la carte. Quand le participant à

localisé les quatre points dans la carte, celle-ci disparaît de l'écran, et réapparaît avec le véhicule positionné sur l'aéroport : les participants peuvent commencer leur trajet.

Le reste de l'expérience se déroule comme dans l'expérience 3.

3.2. Analyse des résultats

Plan d'analyse:

$\underline{S}_{24} * R_6$

Le facteur \underline{S} correspond aux participants ; et le facteur R correspond au rang de l'étape sous-but affectée par le changement dans la résolution de la tâche ($n, n + 1, n - 1, n > 1, n_{final}, NP$).

Les données ont été analysées au moyen d'une procédure GENMOD, et les résultats figurent dans le Tableau 6.

Tableau 6. Score de détection (%) des changements en fonction du rang de l'étape sous-but changée

Rang de l'étape changée					
$n-1$	n	$n+1$	$n > 1$	n_{final}	NP
20.83	62.50	29.17	16.67	16.67	12.50

Les résultats révèlent un effet du rang de l'étape sous-but changée sur la détection des changements ($X^2(5) = 14.08, p < .05$), qui se manifeste par une meilleure performance de détection observée pour les changements de rang n que pour les autres changements (une analyse des contrastes révèle $X^2(1) = 10.53, p < .005$), performance qui diminue au fur et à mesure que l'étape sous-but modifiée s'éloigne du sous-but en cours de résolution.

3.3. Discussion

L'expérience 5 réplique les résultats de l'expérience 3, à savoir que la détection des changements réalisés sur la scène dépend de la tâche : les changements les mieux détectés concernent les étapes qui sont directement nécessaires à la résolution du sous-but dans lequel les participants se sont engagés, ce qui suggère, comme dans l'expérience 3 que l'attention est guidée progressivement vers les informations visuelles de la scène requises par la tâche. Ces informations sont traitées et représentées en mémoire de manière privilégiée, ce qui permet la réussite des processus de comparaisons indispensables à la détection de leur changement. Nous avons formulé, lors de l'analyse des résultats de l'expérience 3, l'hypothèse selon laquelle les détections des changements concernant l'étape but pouvaient résulter de l'utilisation stratégique des panneaux de direction, de type : « si la position de Paris est en désaccord avec les indications de direction, alors il y a eu un changement sur Paris ». Dans l'expérience 5, nous avons effacé de la carte toutes les indications de directions : la détection des changements affectant l'étape but décline considérablement (elle passe de 24.39% dans l'expérience 3 à 16.67% dans l'expérience 5). On peut en conclure que le but de la tâche ne bénéficie pas de traitements visuels et attentionnels particuliers dans cette tâche. A noter également la proportion de changements de rang $n - 1$ détectés, qui révèle l'influence de post-effets attentionnels sur la représentation en mémoire de la scène : la saillance des informations visuelles qui ont été traitées et représentées en mémoire lors de la réalisation d'un sous-but devenu obsolète tend à décliner progressivement avec la progression de la résolution de la tâche.

L'objectif de l'expérience 5 était de distinguer si l'effet de la tâche sur la représentation en mémoire de l'information visuelle était simplement le fait d'une exploration visuelle de la scène restreinte aux éléments visuels pertinents par rapport à la tâche ou si cet effet était

suffisamment robuste pour persister lorsque tous les éléments de la scène avaient préalablement été focalisés par l'œil. En d'autres termes, il s'agissait de déterminer si les représentations visuelles fonctionnelles résultaient d'un simple effet de l'exploration stratégique des scènes, ou si, au contraire, elles se caractérisaient véritablement par une haute sélectivité à l'information visuelle ainsi que par un aspect transitoire. Les résultats montrent que, bien que la sélectivité des représentations visuelles fonctionnelles s'exprime dès la prise d'informations sur la scène, privilégiant ainsi le traitement des éléments de la scène pertinents par rapport à la tâche très tôt dans le traitement visuel, cette sélectivité dans la prise d'information ne peut à elle seule expliquer les effets de la tâche observés sur la représentation de l'information visuelle lorsque les participants doivent réaliser une tâche sur les scènes visuelles.

4. DISCUSSION DU CHAPITRE SIX

L'objectif de ce chapitre était d'étudier la perception de scènes visuelles complexes en condition de vision active et notamment comment les informations visuelles pertinentes par rapport à une tâche sont représentées en mémoire, et quelle est leur évolution en regard de la progression de la tâche. Le principe des expériences présentées dans ce chapitre était de décomposer la tâche à réaliser sur les scènes visuelles en but et sous-buts et de mesurer les détections différentielles des changements réalisés sur chacun des sous-but de la tâche, en considérant les détections des changements comme des indicateurs des éléments de la scène qui ont été focalisés par l'attention et représentés en mémoire. Dans l'ensemble, les résultats obtenus dans les expériences 3, 4, et 5 mettent en évidence le rôle fondamental joué par les buts de la tâche dans la représentation en mémoire d'une scène visuelle complexe. En

particulier, les résultats montrent, tout d'abord, que les participants résolvent la tâche en la décomposant en sous-buts simples de type « atteindre telle étape », ce qui est en accord avec les travaux de Ballard et collaborateurs (1995) qui montrent que, dans une tâche de copie de blocs colorés, les participants tendent à décomposer la tâche en sous-tâches élémentaires de sorte à alléger la charge mnésique requise par la tâche.

Plus spécifiquement, les résultats de ce chapitre se résument en quatre points :

- Les changements les mieux perçus sont ceux qui altèrent les étapes pertinentes par rapport au sous-but en cours de résolution (rangs n et $n+1$), avec une détection préférentielle des changements survenant sur l'étape immédiatement nécessaire à la résolution de ce sous-but, suggérant que ces étapes ont bénéficié d'un traitement attentionnel poussé et ont été préférentiellement représentées en mémoire.
- La détection des changements diminue au fur et à mesure que l'étape sous-but concernée par le changement s'éloigne du sous-but en cours de résolution dans la structure des sous-buts. Ce résultat est intéressant car il indique que la pertinence par rapport à la tâche est une condition nécessaire mais non suffisante à la représentation en mémoire de l'information: le système cognitif traite l'information dont il a besoin au moment où il en a besoin. Cette caractéristique met en exergue l'aspect fonctionnel des représentations visuelles : elles sont construites graduellement par et pour la tâche.
- Une fois qu'un sous-but est satisfait, et que l'attention s'est détournée de l'information visuelle qui lui était associée, la représentation de cette information tend à décliner en mémoire.
- Les étapes non pertinentes par rapport à la tâche sont ignorées des traitements visuels et cognitifs.

Par conséquent, nous pouvons conclure des buts de la tâche qu'ils guident pas à pas la focalisation de l'attention sur les éléments de la scène qui sont pertinents par rapport à la tâche. Ainsi, au fur et à mesure que la résolution progresse, les objets de la scène impliqués dans la résolution de la tâche sont tour à tour focalisés par l'attention et représentés en mémoire de façon suffisamment détaillée pour soutenir la détection consciente de leur changement : la représentation de la scène est, par conséquent dynamique et transitoire, puisqu'elle ne contient que les informations visuelles qui sont directement et immédiatement requises par la résolution du sous-but actuel. La représentation en mémoire d'une même scène diffère donc littéralement selon le sous-but à satisfaire, et il serait vain de considérer qu'il existe une représentation globale de la scène visuelle, qui s'enrichirait du temps passé à son exploration. Là où les résultats des expériences du chapitre précédent suggéraient que la tâche entraîne l'émergence d'un niveau fonctionnel de représentation de la scène, les expériences du présent chapitre révèlent que ce niveau de représentation dépend principalement des exigences du sous-but en cours de résolution. Les représentations visuelles fonctionnelles sont construites par et pour la réalisation de la tâche : elles résultent de l'extraction, du traitement, de la représentation et de l'actualisation sélectives de l'information visuelle au fur et à mesure que celle-ci est requise par les demandes instantanées de la tâche, et ne sauraient uniquement résulter de l'exploration visuelle sélective des scènes.

Le chapitre suivant propose une modélisation des effets de la tâche sur la représentation en mémoire de l'information visuelle.

– CHAPITRE SEPT –

MODÉLISATION DU RÔLE DES BUTS DE LA TÂCHE SUR LA REPRÉSENTATION DE L'INFORMATION VISUELLE EN MÉMOIRE

Les expériences du chapitre précédent ont permis de mettre en évidence plusieurs caractéristiques des représentations visuelles fonctionnelles :

- Un élément de la scène a d'autant plus de chances d'être représenté en mémoire qu'il est pertinent par rapport à la tâche. Les éléments non pertinents par rapport à la tâche tendent, quant à eux, à être ignorés des traitements visuels et cognitifs (voir aussi Gottlieb et al., 1998; Hayhoe, 2000; Land & Hayhoe, 2001; Pelz et al., 2001; Triesch et al., 2003; VanRullen & Koch, 2003; Wallis & Bühlhoff, 2000).
- Un élément de la scène a d'autant plus de chances d'être représenté en mémoire qu'il est pertinent par rapport au sous-but en cours de résolution. La probabilité de représentation en mémoire de l'information visuelle relative à un sous-but de la tâche décroît au fur et à mesure que ce sous-but s'éloigne du sous-but immédiatement requis par la tâche.
- La représentation d'un élément en mémoire tend à décroître quand le sous-but associé à cet élément a été satisfait.

Les représentations visuelles semblent être construites sur un critère fonctionnel, et ne contenir que l'information visuelle immédiatement requise par les demandes de la tâche. Elles sont par conséquent hautement transitoires et dynamiques, et se voient continuellement actualisées au fur et à mesure que la résolution de la tâche progresse.

Le présent chapitre propose une modélisation des représentations visuelles fonctionnelles, qui prend en compte ces caractéristiques. Le modèle que nous proposons ici est un modèle multinomial, aussi appelé modèle MPT⁹⁰. Les modèles multinomiaux sont des modèles statistiques qui permettent de mesurer les processus cognitifs qui sous-tendent les performances observées. Ces modèles, développés exclusivement pour des données catégorielles qui suivent la loi multinomiale⁹¹, supposent que la réalisation d'une tâche repose sur l'atteinte successive de plusieurs états cognitifs, associés à différentes probabilités. Un postulat fondamental de ces modèles est de considérer que ces états cognitifs, qui ne sont pas directement observables, sous-tendent les comportements observés chez les participants. Globalement, les modèles MPT se distinguent par une structure arborescente, dans laquelle chaque branche représente une séquence hypothétique d'étapes de traitement, qui aboutit à une catégorie de réponse, sachant qu'une même catégorie de réponse peut résulter de branches différentes. La modélisation nécessite que le chercheur spécifie avec précision quels sont les états cognitifs supposés être à l'origine du comportement observé, et comment ces différents états interagissent les uns avec les autres. Ces états cognitifs sont représentés, dans le modèle, sous la forme de paramètres qui peuvent être estimés sous la forme de probabilités, à partir des données recueillies par l'expérience. Ainsi, à chaque paramètre p correspond le

⁹⁰ MPT pour: "Multinomial Processing Tree". On peut également les trouver sous l'appellation « modèles polynomiaux »

⁹¹ La loi multinomiale est la généralisation de la loi binomiale à plus de deux catégories de réponses.

paramètre complémentaire ($1 - p$), de sorte que la somme de ces probabilités égale 1. L'adéquation du modèle aux données est ensuite évaluée par des tests d'ajustements (pour une revue, voir Batchelder & Riefer, 1999; et pour le détail technique de la statistique utilisée, voir Hu & Batchelder, 1994).

Notre modèle MPT a pour ambition de déterminer le contenu des représentations visuelles en condition de vision active, c'est-à-dire lorsque les participants doivent réaliser une tâche. Là encore, la représentation en mémoire d'un objet de la scène est estimée par le modèle à partir de la performance de détection des changements, obtenue en couplant une tâche de type résolution de problème à un paradigme de cécité au changement comme ce fut le cas dans les expériences 3, 4, et 5.

1. UN MODÈLE MULTINOMIAL DE LA REPRÉSENTATION DE SCÈNES VISUELLES COMPLEXES EN CONDITION DE VISION ACTIVE

Notre modèle multinomial de la représentation de scènes visuelles complexes en condition de vision active vise à rendre compte du contenu des représentations visuelles lorsque la perception de la scène est dévolue à la satisfaction d'une tâche. Le postulat de base du modèle est que les informations visuelles traitées par l'œil et l'attention ne sont pas forcément représentées en mémoire (Becker & Pashler, 2002; Horowitz & Wolfe, 1998), et que les représentations visuelles de la scène sont dévolues à la résolution de la tâche (Ballard et al., 1995; Hayhoe, 2000; Hayhoe et al., 1998; Hayhoe et al., 2003; Karn & Hayhoe, 2000; Maunsell, 1995; Shinoda et al., 2001; Triesch et al., 2003; Wallis & Bühlhoff, 2000). Par conséquent, le modèle postule que les représentations visuelles contiennent, de manière préférentielle, l'information visuelle qui est directement nécessaire et pertinente à la

réalisation de la tâche. Ainsi, dans le cas d'une tâche pouvant être décomposée en but et sous-but, comme c'est le cas de la tâche utilisée dans les expériences 3, 4, et 5, le modèle postule que les représentations visuelles coderont de manière préférentielle non seulement les informations visuelles pertinentes par rapport à la tâche mais *aussi* pertinentes par rapport au sous-but en cours de résolution. De plus, le modèle considère que le système cognitif fonctionne sur un principe d'économie des ressources cognitives (Ballard et al., 1995; Ballard et al., 1998), et qu'il tendra donc à ne représenter que les informations visuelles qui sont nécessaires et suffisantes à la réalisation de la tâche. En d'autres termes, le modèle postule que les éléments de la scène ne seront représentés que par une seule de leurs propriétés constitutives (identité ou localisation)⁹², si la représentation d'une seule de ces propriétés est suffisante pour réaliser la tâche. Enfin, le modèle prend également en compte le « *visual sensing* », qui définit la capacité de certains participants à pouvoir « sentir » que des changements sont survenus sur une scène visuelle, sans pour autant être capables d'expliciter ces changements (Rensink, 2004).

1.1. Le modèle

Le modèle est conçu pour des tâches avec trois catégories de réponses alternatives (détection correcte du changement, détection partielle du changement, ou non détection du changement), et trois types d'essais (changement d'une étape impliquée dans le sous-but en cours de résolution ; changement d'une étape impliquée dans un autre sous-but ; et changement d'une étape non pertinente par rapport à la tâche). Comme le montre la Figure 15,

⁹² Ces propriétés sont déterminées par les deux voies de traitement visuel : la voie ventrale qui code l'identité de l'information visuelle, et la voie dorsale qui en code la position dans l'espace (Ungerleider & Haxby, 1994; Watanabee, 2003).

le modèle considère qu'un objet⁹³ de la scène peut être représenté en mémoire (R_n) ou non ($1 - R_n$), pertinent par rapport à la tâche (T) ou non ($1 - T$), et pertinent soit par rapport au sous-but en cours de résolution (SBc), soit par rapport à un autre sous-but ($1 - \text{SBc}$). De plus, le modèle spécifie que soit l'une (1P), soit les deux ($1 - 1P$) propriétés constitutives de l'objet (identité *et/ou* localisation) peuvent être représentées en mémoire. Si les deux propriétés de l'objet sont représentées et que cet objet est changé, alors le modèle prédit une détection correcte (DC). Par contre, si une seule des propriétés constitutives de l'objet est représentée, elle peut soit correspondre à la propriété changée (PC ; *par exemple, les participants représentent uniquement la localisation de l'objet, et le changement sur cet objet est un déplacement*), auquel cas le modèle prédit une détection correcte (DC), soit ne pas correspondre à la propriété changée ($1 - \text{PC}$; *par exemple, les participants représentent uniquement l'identité de l'objet, et le changement sur cet objet est un déplacement*). Étant donné que les participants ne savent pas à l'avance quel est le type de changement auquel ils vont être confrontés, nous avons fixé la probabilité de représenter la propriété qui sera changée à la probabilité que donnerait le hasard, c'est-à-dire une chance sur deux ($\text{PC} = .50$). Dans le cas où ils auraient représenté la propriété inchangée, les participants peuvent tout de même avoir la sensation qu'un changement a eu lieu sur la scène (s)⁹⁴ ou non ($1 - s$). S'ils ont la sensation qu'un changement est survenu sur la scène, les participants peuvent deviner (g) ou non ($1 - g$) quel type de changement est survenu. Le choix du type de changement (image en plus, en moins, déplacée, ou différente) conduit à une chance sur quatre de deviner le

⁹³ Dans nos expériences, les objets sont les étapes de la scène.

⁹⁴ Ou « sentir » le changement, en référence au « visual sensing » de Rensink (2004)

changement qui est effectivement survenu sur l'objet (g est donc fixé à $.25$)⁹⁵, et ainsi de détecter correctement le changement (DC). Alternativement, si les participants ne devinent pas le changement (par exemple, si les participants désignent correctement l'étape changée mais se trompent sur le type de changement), alors le modèle prédit une détection partielle (DP). Enfin si les participants ne sentent pas le changement ($1 - s$), ou s'ils n'ont pas représenté l'objet en mémoire ($1 - R$), le modèle prédit une non détection (ND).

⁹⁵ Théoriquement, si les participants ont représenté un objet x , et qu'ils ont la sensation qu'un changement est survenu sur cet objet, on s'attend à ce qu'ils signalent correctement le changement sur cet objet (en choisissant correctement l'étape changée) mais on considère qu'ils peuvent se tromper sur le choix du *type* changement (image en plus ? en moins ? déplacée ? différente ?), ce qui conduit à estimer à une chance sur quatre la probabilité de deviner le changement.

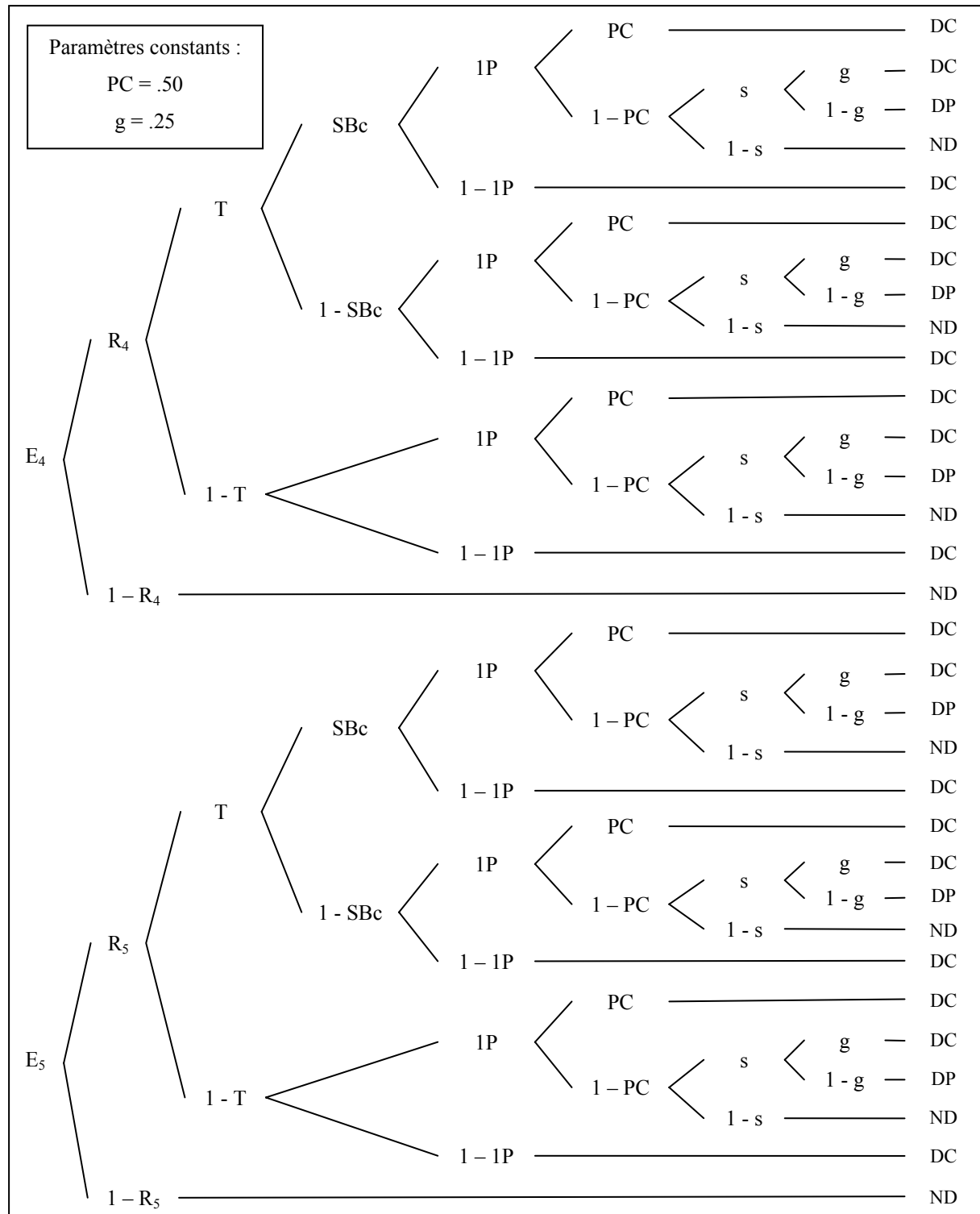


Figure 15. Modèle multinomial de la représentation de scènes visuelles complexes en condition de vision active. E4 et E5 : désigne les expériences 4 et 5 ; R_4 , R_5 = probabilité de représenter un objet de la scène dans les expériences 4 et 5 ; T = probabilité de représenter un élément pertinent par rapport à la tâche ; SBc = probabilité de représenter un élément pertinent par rapport au sous-but en cours de résolution ; 1P = probabilité de ne représenter qu'une seule des propriétés constitutives de l'objet (identité ou localisation) ; PC = probabilité de représenter la propriété changée ; s = probabilité de « sentir » le changement ; g = probabilité de deviner le changement ; DC = détection correcte ; DP = détection partielle ; ND = non-détection

La section suivante propose de tester notre modèle avec les données recueillies dans les expériences 4 et 5⁹⁶.

1.2. Ré-analyse des données des expériences 4 et 5

1.2.1. *Aspects pratiques*

Notre modèle a pour ambition de déterminer le contenu des représentations visuelles en fonction de la tâche, en condition active. Afin de pouvoir traiter conjointement les données dans les conditions indépendantes que représentent les expériences 4 et 5, nous avons utilisé un modèle multinomial joint (voir Figure 15), qui est une catégorie particulière de modèles multinomiaux. Un modèle joint peut être considéré comme un modèle MPT hiérarchique à deux niveaux, un premier niveau reflétant la condition expérimentale (ici, expériences 4 et 5), et un second niveau reflétant les réponses des participants à chaque item (pour une formalisation statistique des modèles joints, voir Hu & Batchelder, 1994).

Comme nous l'avons mentionné plus haut dans le texte, le modèle est conçu pour des tâches dans lesquelles on distingue trois types d'essais : (i) changement sur une étape pertinente par rapport à la tâche et pertinente par rapport au sous-but en cours de résolution, (ii) changement sur une étape pertinente par rapport à la tâche et pertinent par rapport à un autre sous-but que le sous-but en cours de résolution, et (iii) changement non pertinent par

⁹⁶ Nous ne pouvons malheureusement pas tester notre modèle à l'aide des données recueillies dans l'expérience 3, puisque, dans cette expérience, il n'y a pas de changement d'une étape non pertinente par rapport à la tâche.

rapport à la tâche. Nous avons implémenté le modèle avec les données brutes recueillies dans les expériences 4 et 5 comme suit :

- (i) : les changements à la fois pertinents par rapport à la tâche *et* au sous-but en cours de résolution regroupent les données correspondant aux changements de rang n et $n + 1$, c'est-à-dire aux changements de l'étape immédiatement requise par la résolution du sous-but et l'étape suivante⁹⁷
- (ii) : les changements pertinents par rapport à la tâche et pertinents par rapport à un *autre* sous-but que celui en cours de résolution regroupent les données correspondant aux changements de rang $n > 1$ et n_{final}
- (iii) : les changements non pertinents par rapport à la tâche regroupent les données correspondant aux changements NP et rang $n - 1$, puisque les étapes sous-but de rang $n - 1$ ont été antérieurement atteintes et ne sont donc plus pertinentes par rapport à la tâche.

De plus, comme nous en avons fait l'hypothèse dans le chapitre précédent, le démasquage des étapes de la carte, peut, dans l'expérience 4, avoir conduit les participants à réaliser une double tâche sur la carte, double tâche qui aurait conduit à la représentation accrue de toutes les étapes de la carte. L'expérience 5, dans laquelle les étapes ne constituaient qu'un fond « indifférencié » dans lequel il s'agit de trouver une forme, remédie à ce biais potentiel et de fait, ne favorise pas la représentation accrue de toutes les étapes de la carte. Aussi, notre modèle considère que la probabilité de représenter une étape de la carte diffère dans l'expérience 4 et 5, et pose donc $R_4 \neq R_5$. Le modèle suppose par ailleurs que tous les autres paramètres rencontrent la même probabilité d'activation dans les deux expériences.

⁹⁷ Considérons par exemple le trajet suivant : gare \rightarrow cabine \rightarrow entrepôt \rightarrow Imaginons un participant qui se trouve à la gare : le sous-but en cours de résolution est le sous-but « satisfaire la *branche* 'entrepôt' ». L'étape immédiatement impliquée dans la résolution du sous-but (de rang : n) est la cabine, l'étape suivante (de rang : $n+1$) est l'entrepôt

1.2.2. Résultats

Les réponses des participants aux changements des expériences 4 et 5 du chapitre précédent ont été considérées comme des détections correctes (DC) quand les participants rapportaient correctement le type de changement (image en plus, en moins, déplacée, ou différente) *et* l'étape modifiée. S'ils ne rapportaient correctement que l'un *ou* l'autre du type de changement et de l'étape modifiée, la réponse des participants était considérée comme une détection partielle (DP). Enfin une réponse entièrement erronée ou une non réponse étaient considérées comme des non-détections (ND). Les nombres de détections correctes, détections partielles et non-détections, sommés pour tous les participants et tous les essais, figurent en annexe 5.

Un des avantages des modèles multinomiaux est de donner une estimation de chacun des paramètres du modèle, exprimés sous la forme de probabilités. Ces probabilités sont calculées à partir d'équations établies à partir des probabilités de réponses, qui sont elles-mêmes calculées pour chaque catégorie de réponse (DC, DP, ND). La probabilité pour une certaine catégorie de réponse résulte de la sommation des branches de l'arbre de traitement qui conduisent à cette réponse. L'estimation statistique de la probabilité d'activations des paramètres par le modèle MPT nécessite que le modèle soit globalement identifiable⁹⁸ (Batchelder & Riefer, 1999). Un modèle est globalement identifiable si ses paramètres peuvent être exprimés uniquement en terme de probabilités de réponse (Erdfelder, sous presse). Tel quel, le modèle présenté en Figure 15 n'est pas globalement identifiable. Une manière de remédier aux problèmes d'identification des modèles consiste à attribuer des valeurs numériques à certains paramètres (Batchelder & Riefer, 1999). Nous avons donc imposé une valeur théorique à l'un de nos paramètres, et avons attribué au paramètre s la

⁹⁸ Notons que l'identifiabilité globale n'est pas nécessaire pour tester l'ajustement du modèle aux données

valeur .33, sur la base d'arguments théoriques (Rensink, 2000b, 2004)⁹⁹. Ainsi contraint, le modèle est globalement identifiable (pour un exemple de l'identification d'un modèle MPT, voir annexe 10).

A partir des données des expériences 4 et 5, nous pouvons donc obtenir une évaluation de chacun des paramètres du modèle par l'estimation d'un coefficient maximum de vraisemblance et un test de khi-deux d'ajustement, qui suit la distribution khi-deux (Hu & Batchelder, 1994), et, pour chaque condition expérimentale, à l'aide du programme informatique GPT.exe (Hu, 1999).

Nous avons fixé le risque alpha à .05 pour toutes les analyses statistiques rapportées dans cette section.

Avec $N = 288$ (48 participants * 6 essais), sept degrés de liberté (dl), et un risque de première espèce $\alpha = .05$, un test khi-deux d'ajustement indique une valeur $X_0^2(7) = 2.88$. Cette valeur est très inférieure à la valeur critique $VC = 14.07$ donnée par la table de khi-deux pour 7 dl, ce qui indique que notre modèle tel que l'illustre la Figure 15 est tout à fait en mesure de rendre compte des données observées dans les expériences 4 et 5. Plus précisément, ce résultat indique que les paramètres du modèle tels qu'ils ont été pensés, ainsi que les contraintes apportées au modèle (probabilité de représenter la propriété sur laquelle porte le changement de .5, probabilité de deviner le changement de .25 et probabilité de sentir le changement de .33), sont validés par les données empiriques. Ce résultat est capital car il indique que le modèle présenté en Figure 15 est tout à fait à même d'expliquer les comportements mesurés chez des participants humains en condition de vision active.

⁹⁹ En effet, Rensink (2000b, 2004) estime à un tiers la proportion de participants capables de manifester la sensation d'un changement sans pouvoir l'expérimenter consciemment.

Un aspect fondamental de notre modèle est de considérer que les objets visuels ont d'autant plus de chances d'être représentés en mémoire qu'ils sont pertinents par rapport à la tâche (T) et pertinents par rapport au sous-but en cours de résolution (SBc). Testons cet aspect fonctionnel de nos représentations visuelles : si l'on suppose que nos représentations visuelles ne répondent pas à un critère fonctionnel tel que nous l'avons défini dans les chapitres précédents, alors l'on considère que ni la pertinence par rapport à la tâche, ni la pertinence par rapport au sous-but en cours de résolution n'influencent le contenu des représentations visuelles. Cette supposition revient à forcer le modèle à considérer qu'un objet visuel a autant de chances d'être représenté en mémoire qu'il soit ou non pertinent par rapport à la tâche, ce qui nous conduit à poser la contrainte : $T = 1 - T$. Avec cette contrainte, le test d'adéquation du modèle aux données indique $X_1^2(8) = 24.55$, avec 8 degrés de liberté. Cette valeur est supérieure à la valeur critique $VC = 15.51$, donnée par la table de distribution du khi-deux pour 8 dl, ce qui indique que le modèle contraint à considérer que la pertinence par rapport à la tâche n'influence pas les représentations visuelles n'est plus en mesure de rendre compte des données observées. Ce premier résultat montre clairement que la pertinence par rapport à la tâche est un facteur fondamental de la représentation des scènes visuelles en mémoire.

Qu'en est-il de l'effet de l'immédiateté des exigences des sous-buts ? Si on suppose que la pertinence par rapport aux exigences du sous-but en cours de résolution n'est pas un facteur crucial de la représentation de l'objet en mémoire, alors on contraint le modèle à considérer qu'un objet visuel a autant de chances d'être représenté en mémoire qu'il soit impliqué ou non dans la résolution du sous-but en cours, ce qui se traduit par : $SBc = 1 - SBc$. On observe que l'ajustement du modèle ainsi contraint aux données est estimé à $X_2^2(8) = 10.36$ avec 8 dl, valeur inférieure à la valeur critique indiquée pour $VC = 15.51$. Ce résultat suggère que le modèle ainsi contraint parvient tout de même à rendre compte des données observées dans

l'expérience. Afin de déterminer lequel du modèle initial ou du modèle contraint permet de simuler le plus finement les données empiriques, il est nécessaire de calculer un indice ΔX^2 tel que $\Delta X^2 = X_2^2 - X_0^2$, assorti d'un nombre de degrés de liberté Δdl tel que $\Delta d = dl_2 - dl_0$, qui va nous indiquer dans quelle mesure la contrainte nouvellement apportée modifie le modèle et rend compte des données. Un ΔX^2 qui serait supérieur à la valeur critique donnée pour Δdl degrés de liberté signifierait que le modèle contraint diffère du modèle initial, et qu'il vaut mieux accepter comme étant le modèle s'ajustant le mieux aux données le modèle assorti du moins de degré de liberté¹⁰⁰. Par contre, un ΔX^2 qui serait inférieur à la valeur critique donnée pour Δdl degrés de liberté signifierait que le modèle contraint ne diffère pas du modèle initial, et qu'il est plus sage de choisir comme « meilleur » modèle le modèle le plus général, c'est-à-dire le modèle assorti du plus grand nombre de degrés de liberté. Notre calcul du ΔX^2 nous donne $\Delta X^2 = 7.48$ et $\Delta dl = 1$, ce qui indique, avec un risque d'erreur $\alpha = .05$ et une valeur critique $VC = 3.84$, que le modèle contraint se distingue du modèle initial, et qu'il faut garder le modèle initial comme étant le modèle qui permet de rendre compte des données de la manière la plus fine. Il en résulte que l'immédiateté des exigences de la tâche est bel et bien un facteur déterminant dans la représentation en mémoire des objets visuels.

Enfin, si l'on force le modèle à considérer qu'un objet visuel a autant de chances d'être représenté en mémoire qu'il soit ou non pertinent par rapport à la tâche *et* qu'il soit ou non pertinent par rapport au sous-but en cours (ce qui nous conduit à poser les contraintes $T = 1 - T$ *et* $SBC = 1 - SBC$), l'évaluation de l'ajustement du modèle aux données est de $X_3^2(9) = 32.03$, avec 9 degrés de liberté. Cette valeur est bien supérieure à la valeur critique $VC = 16.92$, ce

¹⁰⁰ Un modèle qui s'ajusterait tout à fait aux données aurait zéro degrés de liberté, on parle dans ce cas de modèle saturé

qui indique qu'ainsi contraint, le modèle n'est plus du tout en mesure de rendre compte des données recueillies dans les expériences 4 et 5.

Le modèle le plus à même de prédire et d'expliquer les données observées dans les expériences 4 et 5, est donc bien le modèle présenté dans la Figure 15. Ainsi, notre modèle témoigne de l'aspect capital joué par l'immédiateté des demandes de la tâche sur la représentation visuelle d'une scène naturelle complexe.

Comme nous l'avons mentionné plus haut, une des caractéristiques majeures des modèles MPT est de proposer une estimation de chacun des paramètres du modèle, à partir des données empiriques, exprimée sous la forme de probabilités. Le Tableau 7 fournit les estimations des paramètres de notre modèle multinomial de la représentation de scènes visuelles complexes en condition de vision active illustré en Figure 15, à partir des données des expériences 4 et 5.

Tableau 7. Estimations des paramètres du modèle multinomial de la représentation de scènes visuelles complexes en condition de vision active, pour les expériences 4 et 5. R_4 , R_5 = probabilité de représenter un objet de la scène dans les expériences 4 et 5 ; T = probabilité de représenter un élément pertinent par rapport à la tâche ; SBc = probabilité de représenter un élément pertinent par rapport au sous-but en cours de résolution ; $1P$ = probabilité de ne représenter qu'une seule des propriétés constitutives de l'objet (identité ou localisation) ; PC = probabilité de représenter la propriété changée ; s = probabilité de « sentir » le changement ; g = probabilité de deviner le changement ; * = maintenu constant. Déviations standard entre parenthèses. Les équations à l'origine de l'estimation des paramètres sont consultables en annexe 6.

Paramètre du modèle							
R_4	R_5	T	SBc	$1P$	PC	s	g
.39	.36	.74	.66	.53	.50*	.33*	.25*
(.05)	(.05)	(.04)	(.06)	(.16)			

Le modèle montre que les consignes de l'expérience 4 ont conduit les participants à représenter légèrement plus d'objets de la scène visuelle que les consignes de l'expérience 5

(le modèle estime en effet qu'un objet de la scène a 39% de chances d'être représenté dans l'expérience 4, contre 36% dans l'expérience 5). De plus, le modèle estime, conformément à nos attentes, à 74% et 66% de chances qu'un objet de la scène soit représenté s'il est pertinent par rapport à la tâche, et s'il est impliqué dans le sous-but en cours de résolution. Le modèle estime à 53% les chances de ne représenter qu'une des deux propriétés constitutives des objets (identité et localisation), et nous avons fixé à 50% (c'est-à-dire au hasard) les chances que les participants se représentent la propriété sur laquelle porte le changement, puisque les participants ne savent pas à l'avance à quels types de changement ils vont être confrontés. Enfin, nous avons également fixé à 33% la probabilité que les participants aient la sensation d'un changement, et à 25% la probabilité qu'ils devinent quel changement est survenu sur les étapes.

1.3. Discussion

Notre modèle multinomial permet de rendre compte de l'évolution des représentations visuelles des scènes complexes en condition de vision active, c'est-à-dire lorsque le traitement visuel de la scène est dévolu à la réalisation d'une tâche. Le modèle témoigne du caractère fonctionnel de nos représentations visuelles en condition de vision active, qui n'offrent à aucun moment une description détaillée de l'ensemble de la scène. Au contraire, ne sont représentées que les informations visuelles qui sont directement et immédiatement requises par la tâche, le reste des informations de la scène tendant à ne pas être représenté en mémoire même si ces dernières demeurent pertinentes pour la poursuite de la tâche, c'est-à-dire pour la résolution du but. Les représentations visuelles fonctionnelles sont construites par et pour la résolution de la tâche, avec priorité représentationnelle accordée aux éléments visuels les plus pertinents à l'instant t , ce qui est conforme avec les limitations attentionnelles (Irwin, 1992;

Irwin & Zelinsky, 2002; Luck & Vogel, 1997; Rensink, 2000c; Vogel et al., 2001). Ainsi, le modèle montre que nos représentations visuelles sont en perpétuelle restructuration, et proposent, à chaque instant, des descriptions transitoires et particularisées de la scène visuelle qui s'offre à nous. Enfin, le modèle estime que les objets de la scène ont autant de chances d'être représentés selon l'une de leurs propriétés constitutives que selon les deux. Ce résultat mérite que l'on y réfléchisse, puisqu'il semble suggérer qu'il y a autant de chances que le système cognitif fonctionne à l'économie en ne représentant que partiellement les objets de la scène (par une seule de leurs propriétés constitutives, localisation *ou* identité), ou qu'il code, au contraire, très précisément les objets de la scène (par leurs deux propriétés constitutives, localisation *et* identité).

Une explication pourrait provenir des expériences en elles-mêmes plutôt que des caractéristiques du système cognitif. En effet, dans les expériences 4 et 5 on ne procédait qu'à un seul type de changement sur les scènes (des déplacements) et il est donc possible qu'il y ait eu une sorte d'apprentissage inconscient de cette régularité de l'expérience par les participants (Chun & Jiang, 1998), qui en ont ensuite stratégiquement tiré parti pour ne mémoriser qu'une des propriétés des objets de la scène (*eg., il suffit de ne traiter que la localisation des étapes pour en détecter le changement*)¹⁰¹. Cette stratégie serait inapplicable dans une expérience dans laquelle plusieurs types de changements seraient introduits dans la scène. Si une telle expérience répliquait la probabilité équivalente de ne représenter qu'une seule ou les deux des propriétés constitutives des objets visuels, alors il faudrait écarter l'idée que le système cognitif fonctionne à l'économie et tend à ne représenter que l'information d'une scène directement nécessaire et juste suffisante pour réaliser la tâche. Ainsi, il faudrait considérer

¹⁰¹ Rappelons que les deux paramètres (PC et g) qui auraient dû être affectés par une éventuelle stratégie de type « ne coder que la localisation des objets » ont été maintenus constants. Il est donc possible que cette stratégie se soit en fait répercutée sur le paramètre libre situé juste en amont des paramètres PC et g, à savoir le paramètre 1P qui correspond à la probabilité de ne représenter qu'une seule ou les deux propriétés de l'objet.

que certainement d'autres facteurs influencent le nombre de propriétés de l'objet qui seront représentées en mémoire (par exemple, le bruit représenté par la complexité ou le nombre de distracteurs) (Wilken & Ma, 2004).

La section suivante teste cette hypothèse et explore plus en détail l'influence de la tâche sur la représentation des objets visuels et de leurs propriétés en mémoire.

2. HYPOTHÈSES SUR L'INFLUENCE DE LA TÂCHE SUR LA REPRÉSENTATION DES OBJETS VISUELS ET DE LEURS PROPRIÉTÉS CONSTITUTIVES

L'objectif de cette section est de tester si les connaissances concernant la tâche peuvent renforcer le caractère fonctionnel des représentations visuelles en influençant notamment le degré de détail avec lequel les objets de la scène sont représentés. En effet, bien que les expériences des chapitres précédents aient montré que, en condition de vision active, un niveau de représentation fonctionnel de la scène émerge et code spécifiquement les informations visuelles directement nécessaires et pertinentes par rapport à la tâche, elles ne permettent pas de déterminer avec quelle précision les objets de la scène sont représentés. Le présent chapitre propose de tester plus finement les caractéristiques fonctionnelles des représentations visuelles et de déterminer si le principe d'économie cognitive conduit à la représentation des propriétés des objets de la scène nécessaires et suffisantes à la réalisation de la tâche.

Nous savons que les informations visuelles contenues au sein du niveau fonctionnel des représentations visuelles sont transitoires et ne concernent que les éléments de la scène qui sont directement et immédiatement impliqués dans le sous-but en cours de résolution. Nous supposons par ailleurs que les informations codées par les représentations visuelles fonctionnelles sont juste assez détaillées pour satisfaire la résolution du sous-but courant (Ballard et al., 1995). Si tel est le cas, la représentation visuelle d'une même scène consécutivement à la réalisation d'une même tâche devrait dépendre de la façon dont les participants vont se représenter la tâche, et donc du point de vue qu'ils vont adopter pour la résoudre.

Appliquée à la représentation des objets de la scène, cette hypothèse nous amène à considérer qu'un point de vue α orientant l'attention vers telle propriété des objets, et un point de vue β orientant l'attention vers telle autre propriété des objets, devraient conduire à *deux* représentations visuelles fonctionnelles *différentes* d'un *même* objet d'une scène, bien qu'il s'agisse de la même tâche. Concrètement, notre hypothèse est que, si on oriente l'attention des participants vers l'une des deux propriétés constitutives (localisation ou identité) de l'objet plutôt que sur l'autre en manipulant leur point de vue sur la tâche, on devrait observer que la représentation de cet objet diffère selon le point de vue adopté (localisation ou identité). Par exemple, si le système cognitif ne représente que l'information juste suffisante pour satisfaire la tâche (comme nous en faisons l'hypothèse) et qu'une consigne demande aux participants de faire attention à la localisation des étapes (point de vue alors adopté : « localisation »), on s'attend à ce que les déplacements d'étapes (qui constituent des changements cohérents avec le point de vue puisqu'ils altèrent la propriété « localisation » des étapes) soient mieux détectés que les changements d'identité (qui, eux, constituent des changements incohérents avec le point de vue puisqu'ils altèrent la propriété « identité » des étapes). Par contre, si le système cognitif représente les objets pertinents par rapport à la tâche

de manière détaillée (en codant leur identité *et* leur localisation), alors les deux types de changements (localisation et identité) devraient être détectés quel que soit le point de vue adopté.

L'hypothèse testée dans cette section est donc qu'une consigne mettant l'accent sur le fait de traiter une propriété plutôt qu'une autre (par exemple, une consigne qui indique que traiter la localisation des étapes est capital pour satisfaire la tâche) devrait conduire le système cognitif à traiter exclusivement cette propriété. Étant donné que les objets de la tâche qui sont les plus susceptibles d'être représentés en mémoire sont immédiatement et directement nécessaires à la résolution de la tâche à l'instant t , nous supposons que l'effet de la consigne ne devrait réellement s'exercer que sur ces objets pertinents par rapport au sous-but en cours de résolution¹⁰².

Ces hypothèses sont testées avec les données d'une expérience du même type que les expériences du chapitre précédent.

Dans cette expérience, le point de vue des participants est modulé par une consigne qui souligne l'importance de prêter attention à l'une ou l'autre des propriétés constitutives des étapes de la scène (localisation ou identité), et les participants sont confrontés à trois types de changement (déplacement, disparition ou changement d'identité d'une étape). Le changement réalisé sur les scènes est soit cohérent, soit incohérent avec le point de vue adopté. Un changement cohérent avec le point de vue adopté altère la propriété de l'objet accentuée par le

¹⁰² L'hypothèse testée dans cette section n'est pas sans rappeler les hypothèses à l'origine des réseaux sémantiques d'action, qui sont des modèles de description des savoir-faire. Ces modèles montrent que, parmi toutes les propriétés des objets, les propriétés auxquelles on accède directement sont les propriétés fonctionnelles, qui sont activées par les buts que les objets permettent d'atteindre. Par exemple, lorsque que l'on doit « rebooter » manuellement un ordinateur qui a bogué, ce qui consiste à satisfaire la tâche « appuyer sur un petit bouton situé dans un renfoncement », on accède directement à la propriété fonctionnelle d'un stylo (objet à pointe fine) requise par la tâche, alors que cet objet se définit pourtant par d'autres propriétés plus manifestes (objet de forme longiligne, qui sert à écrire...) (Poitrenaud, 1995). Notre hypothèse partage l'idée que l'on accède directement aux propriétés fonctionnelles des objets de la scène, ces propriétés étant dépendantes du point de vue sur la tâche (voir aussi Bastien, 1997).

point de vue (par exemple, le point de vue met l'accent sur la propriété « localisation » de l'objet et le changement est un déplacement), alors qu'un changement incohérent avec le point de vue adopté altère l'autre propriété de l'objet que celle accentuée par le point de vue (par exemple, le point de vue met l'accent sur la propriété « localisation » de l'objet et le changement est un changement d'identité). On recueille les réponses au changement (correctes, partielles et incorrectes).

3. EXPÉRIENCE 6 : INFLUENCE DU POINT DE VUE ADOPTÉ POUR RÉSOUTRE LA TÂCHE SUR LA REPRÉSENTATION DES OBJETS VISUELS ET DE LEURS PROPRIÉTÉS CONSTITUTIVES

Dans cette expérience, il s'agit de manipuler le point de vue des participants sur les propriétés constitutives des objets visuels et d'en observer les effets sur la détection des changements, en fonction du rang de l'étape changée dans la structure des sous-buts.

3.1. Méthode

3.1.1. Participants

Cent trente-deux étudiants inscrits en licence de Psychologie de l'Université de Provence ont pris part à cette expérience. Les participants ont été aléatoirement répartis en 6 groupes de 22 participants. Tous ont attesté d'une acuité visuelle normale ou corrigée et

étaient naïfs du point de vue des objectifs de cette recherche. Aucun d'entre eux n'a participé aux expériences précédentes.

3.1.2. Matériel

Les cartes

Les cartes utilisées sont similaires à celles des expériences 4 et 5, et ne contiennent donc aucune signalisation de direction. La nouveauté de l'expérience 6 est de proposer deux icônes différentes pour chaque étape. La Figure 16 illustre les différentes icônes qui peuvent être rencontrées dans les cartes. La présence de l'une ou l'autre des icônes de chaque étape est contrebalancée dans les 6 cartes.



Figure 16. Illustration des deux types d'icônes que l'on peut rencontrer dans les cartes

Les écrans

L'écran de signalisation des changements permet aux participants de désigner le changement qu'ils ont perçu sur la scène, en indiquant quel type de changement ils ont perçu

et quelle était l'étape altérée par ce changement. Cet écran propose, comme dans les expériences 3, 4, et 5, quatre items sur le type de changements (image en plus, image en moins, image déplacée et image différente), ainsi qu'un cinquième item qui permet aux participants d'indiquer qu'ils ne savent pas de quel type de changement il s'est agi (je ne sais pas). Cet écran propose, de plus dans l'expérience 6, systématiquement les deux icônes de chaque étape.

L'écran d'affichage de la consigne, qui incite, quant à lui, les participants à orienter leur point de vue vers telle ou telle propriété des étapes de la scène visuelle, précède l'affichage de chaque carte dans laquelle les participants auront à déplacer leur véhicule. Cette consigne enjoint aux participants de « Surveiller la localisation des étapes » (le texte apparaît au centre de l'écran, en police jaune sur fond bleu) ou de « Surveiller l'image des étapes » (le texte apparaît au centre de l'écran, en police jaune sur fond rouge). L'écran disparaît quand les participants cliquent sur un bouton « commencer ».

Les arbres des tâches

Comme dans les expériences 3, 4, et 5, le trajet à effectuer concerne 7 des 13 étapes de la carte. Ces étapes sont répertoriées dans un arbre des tâches. La nouveauté de l'expérience 6 est de proposer un arbre des tâches différent pour chacun des 6 groupes de participants, afin de permettre que chacune des six étapes de l'arbre (la gare, l'entrepôt, l'église, la préfecture, Paris et le kiosque), qui fait l'objet d'un changement, soit impliquée dans chacun des six rangs de sous-but de la tâche (sous-but de rang : n ; $n + 1$; $n - 1$; $n > 1$; n_{final} ; et NP). Les six arbres des tâches sont illustrés en annexe 7. Ces six arbres des tâches ont la même structure que l'arbre des tâches des expériences 3, 4, et 5, et ne diffèrent les uns des autres que par le niveau occupé par les étapes qu'ils proposent (une étape super-ordonnée dans un des arbres sera une étape sous-ordonnée dans un autre arbre etc...), de telle sorte que, tous

groupes confondus, toutes les étapes changées aient été impliquées dans chacun des rangs de changement manipulés (n , $n + 1$; $n - 1$; $n > 1$; n_{final} ; NP). La Figure 17 illustre deux des arbres des tâches utilisés dans l'expérience 6.

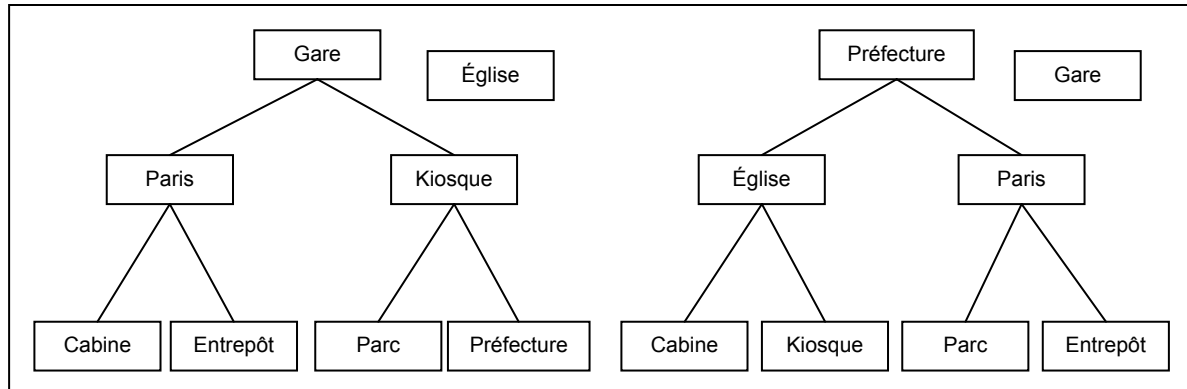


Figure 17. Exemples de deux arbres de tâches utilisés dans l'expérience 6. Comme on peut le constater, les deux arbres mettent en scène les mêmes étapes, mais à des niveaux différents (par exemple, la gare fait office d'étape but dans l'arbre de gauche, et d'étape non pertinente dans l'arbre de droite).

Les règles qui définissent les arbres des tâches (on ne peut atteindre une étape que si les étapes qui lui sont sous-ordonnées ont été atteintes au préalable etc...) sont identiques aux règles de l'arbre des tâches des expériences 3, 4, et 5. Comme dans les expériences précédentes, l'arbre des tâches reste inchangé pour les six cartes : chaque groupe ne voit donc qu'un seul arbre des tâches.

Les changements

L'expérience 6 propose trois types de changements : des déplacements d'étapes d'une position pré- à une position post-changement qui était jusqu'alors inoccupée, des disparitions d'étapes, et des changements d'identité qui consistent en la substitution de l'icône de l'étape par l'autre icône de cette étape (voir Figure 16).

Dans l'expérience 6, comme dans les expériences précédentes, il n'y a qu'un changement par carte (et donc par trajet réalisé). Le changement introduit dans chaque carte

affecte une étape impliquée dans différents sous-buts (de rang n ; $n + 1$; $n - 1$; $n > 1$; n_{final} ; ou NP).

Les six étapes changées sont l'église, la préfecture, l'entrepôt, Paris, le kiosque et la gare, et le rang de chacune de ces étapes dans la résolution de la tâche est contrebalancé dans les six groupes : par exemple, quand le changement de l'entrepôt correspondra au changement de rang n pour un groupe, il correspondra au changement de rang $n > 1$ pour un autre groupe etc... Aussi, dans l'expérience six, chacune des étapes changées est impliquée dans chacun des rangs du changement de sorte à contrôler la saillance perceptive des changements¹⁰³. La répartition de chaque étape changée selon le rang du changement et le groupe est déterminée par un plan en carré latin, consultable en annexe 8.

Enfin, les changements introduits dans les cartes peuvent être soit cohérents soit incohérents avec le point de vue manipulé dans l'expérience. Un changement de type « déplacement » (qui altère la propriété « localisation » des étapes de la scène) est cohérent avec un point de vue accentuant la propriété « localisation » des étapes¹⁰⁴, et incohérent avec un point de vue accentuant la propriété « identité » des étapes¹⁰⁵. Un changement de type

¹⁰³ En effet, il est possible, bien que peu probable, que les résultats des expériences 4 et 5 du chapitre précédent, qui ont servi à implémenter le modèle, ne résultent que d'un effet de matériel. En effet, dans ces expériences, nous avons mesuré la détection des changements en fonction du rang de l'étape changée dans la résolution de la tâche. Pour cela, nous avons fait correspondre à chaque carte un rang de changement, qui se matérialisait par le changement d'une étape particulière de la scène. Par exemple, sur une carte γ , le changement était de rang n et survenait sur l'étape « entrepôt », alors que sur une carte δ , le changement était de rang $n > 1$ et survenait sur l'étape « préfecture ». Il est possible que les meilleures performances de détections du changement de l'entrepôt par rapport à la celui de la préfecture que nous avons attribuées à un effet de la tâche résultent simplement d'un effet de la saillance perceptive : le changement de l'entrepôt sur la carte γ aurait été plus facile à percevoir que le celui de la préfecture sur la carte δ . Si tel était le cas, toutes les conclusions tirées à partir des résultats du chapitre précédent seraient à revoir, et rendraient caduc notre modèle. L'expérience 6 pallie cette critique en faisant en sorte que l'étape changée dans chaque carte soit impliquée dans chacun des rangs des changements, de sorte qu'un même changement (et donc une même saillance perceptive) survient à différents moments de la résolution. Si l'expérience 6 réplique l'effet de la tâche sur la détection des changements, alors elle permettra d'asseoir le rôle fondamental de la tâche dans la perception de scènes visuelles, et de rejeter définitivement l'alternative perceptive dans l'interprétation de nos résultats. Dans le cas contraire, nous devrions admettre qu'il serait prudent de faire moins de cas de l'influence de la tâche sur la perception de scènes visuelles complexes

¹⁰⁴ « Surveillez la position des étapes »

¹⁰⁵ « Surveillez l'image des étapes »

« identité » (qui altère la propriété « identité » des étapes de la scène) est cohérent avec un point de vue accentuant la propriété « identité » des étapes, et incohérent avec un point de vue accentuant la propriété « localisation » des étapes. Enfin, un changement de type « disparition d'étape » est considéré comme cohérent avec un point de vue accentuant la propriété « identité » des étapes, et incohérent avec un point de vue accentuant la propriété « localisation » des étapes¹⁰⁶.

Le point de vue est en accord avec les changements réalisés sur trois des six cartes, et en désaccord avec les changements réalisés sur les trois autres cartes : il y a donc trois changements différents (un déplacement, une disparition et un changement d'identité) par condition de cohérence (cohérent vs. incohérent) entre le point de vue et le changement.

3.1.3. Équipement

L'équipement est le même que dans les expériences 3, 4, et 5.

3.1.4. Procédure

La passation de l'expérience est très proche de celle de l'expérience 3, 4, et 5 (voir Figure 12).

Un nouvel essai commence avec la présentation d'un écran de transition, qui disparaît lorsque les participants cliquent sur un bouton « Prêt ».

¹⁰⁶ Ce choix est déterminé par le raisonnement suivant : si les participants ont porté attention à l'identité d'une étape et que cette étape disparaît de la carte, alors les participants sont en mesure de signaler quel est le type de changement perçu (une « image en moins ») et de choisir correctement quelle était l'icône de l'étape disparue, entre les deux icônes proposées pour cette étape. Par contre, si les participants ont porté attention à la localisation d'une étape et que cette étape disparaît de la carte, alors les participants sont toujours en mesure de signaler quel est le type de changement perçu (une « image en moins »), mais le choix de l'icône de l'étape changée devient plus ardu. Aussi, nous considérerons qu'une consigne accentuant l'identité plutôt que la localisation des étapes sera cohérente avec les changements de type « disparition d'étape ».

Avant la présentation de chaque carte, une consigne¹⁰⁷ s'affiche au milieu de l'écran. Cette consigne enjoint aux participants de porter attention soit à la localisation des étapes dans la carte, soit à l'image des étapes dans la carte. Quand les participants cliquent sur un bouton « commencer », cet écran disparaît et s'affiche alors la première carte, dans laquelle le véhicule à déplacer est situé à l'aéroport. Le reste de l'expérience est identique aux expériences du chapitre précédent.

Avant de commencer l'expérience, les participants passent une phase d'entraînement, sur une carte simple, différente de celles utilisées dans l'expérience. La carte de l'entraînement est la même que celle utilisée dans les expériences 3, 4, et 5. Cet entraînement se déroule en trois temps comme dans les expériences du chapitre précédent. Dans un premier temps, l'expérimentateur fait une démonstration de tous les déplacements (licites et illicites) dans la carte avec le véhicule, et de la manière dont il faut signaler les changements, ensuite les participants réalisent eux même deux fois le trajet en signalant les changements. A l'issue de l'entraînement, les participants ont été confrontés aux trois types de changements de l'expérience.

¹⁰⁷ « Surveillez la localisation des étapes » ou « Surveillez l'image des étapes » selon le point de vue correspondant à la carte (voir annexe 8)

4. MODÉLISATION DE L'INFLUENCE DU POINT DE VUE SUR LA REPRÉSENTATION DES SCÈNES VISUELLES

4.1. Les prédictions de notre modèle multinomial

Notre modèle de la représentation de scènes visuelles complexes en condition de vision active devrait nous permettre de rendre compte de l'effet du point de vue adopté sur la représentation des propriétés des objets visuels. Nous rappelons que le modèle suppose que les représentations visuelles sont dévolues à la réalisation de la tâche, et qu'elles codent les objets de la scène pertinents par rapport à la tâche et pertinents par rapport au sous-but en cours de traitement. Le modèle suppose de plus que, dans un souci d'économie cognitive, les objets visuels ne seront pas représentés dans leur détail, et que le système tendra à ne représenter que l'une des deux propriétés constitutives des objets (localisation ou identité), si la représentation d'une de ces deux propriétés seulement suffit à la réalisation de la tâche. Enfin, le modèle prend également en compte le « *visual sensing* », et la possibilité que les participants puissent deviner quel changement a eu lieu sur les scènes. Il est à noter que la proportion de « *visual sensing* » sera laissée libre, afin de permettre au modèle, qui est globalement identifiable ainsi, d'estimer la proportion de réponses dues à une sensation du changement.

L'enjeu de cette section est d'utiliser notre modèle multinomial afin de montrer que la représentation des objets visuels va dépendre du point de vue adopté pour résoudre la tâche. En effet, on attend du point de vue, qui met l'accent sur une des propriétés des objets de la scène, qu'il influence la probabilité, estimée par le modèle, de représenter la propriété

changée. Ainsi, la probabilité de représenter la propriété changée devrait augmenter lorsque cette propriété a été préalablement accentuée par la consigne (condition : changement cohérent avec le point de vue) et diminuer lorsqu'elle n'a pas été préalablement accentuée par la consigne (condition : changement incohérent avec le point de vue). Plus précisément, l'hypothèse que nous formulons est, qu'afin de minimiser la charge cognitive de la tâche en mémoire, le système cognitif ne devrait représenter que la quantité d'information directement nécessaire et juste suffisante pour réaliser la tâche (Ballard et al., 1995), et que, par conséquent le point de vue ne devrait réellement influencer *que* la représentation des objets visuels pertinents par rapport à la tâche *et* pertinents par rapport au sous-but en cours de résolution.

Comme le montre la Figure 18, le premier niveau du modèle joint reflète la condition expérimentale : le changement est soit cohérent (c) , soit incohérent (i) avec le point de vue manipulé, et le second niveau donne les performances observées pour chaque item de chaque condition.

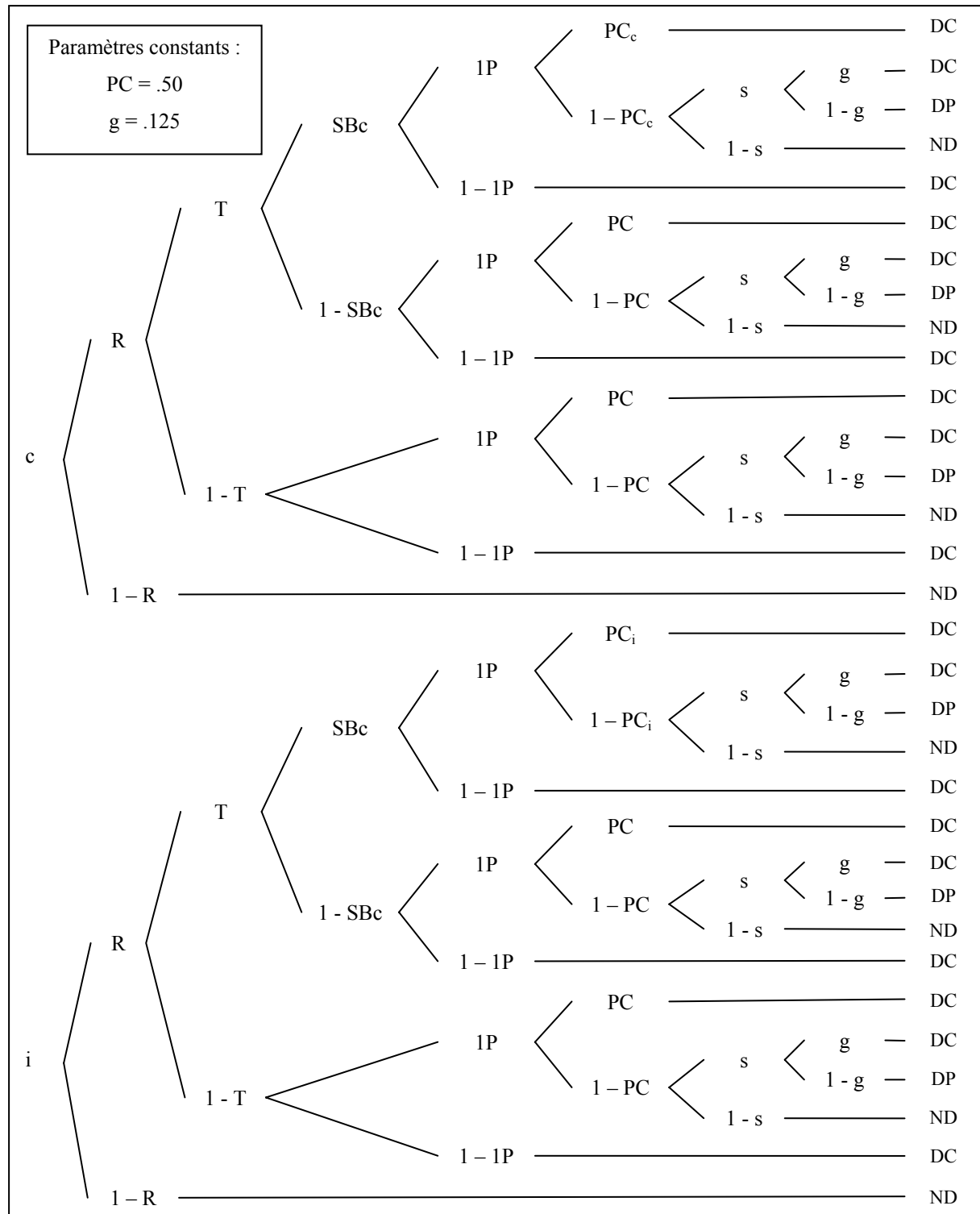


Figure 18. Modèle multinomial de la représentation de scènes visuelles complexes en condition de vision active. c et i désignent les conditions du changement, selon que le changement est cohérent (c) ou incohérent (i) avec le point de vue manipulé. R = probabilité de représenter un objet de la scène; T = probabilité de représenter un élément pertinent par rapport à la tâche; SBc = probabilité de représenter un élément pertinent par rapport au sous-but en cours de résolution; 1P = probabilité de ne représenter qu'une seule des propriétés constitutives de l'objet (identité ou localisation); PC = probabilité de représenter la propriété changée; PC_c = probabilité de représenter la propriété changée dans la condition « changement cohérent avec le point de vue »; PC_i = probabilité de représenter la propriété changée dans la condition « changement incohérent avec le point de vue »; s = probabilité de « sentir » le changement; g = probabilité de deviner le changement; DC = détection correcte; DP = détection partielle; ND = non-détection.

Comme précédemment, le modèle considère qu'un objet de la scène peut être représenté en mémoire (R) ou non ($1 - R$), pertinent par rapport à la tâche (T) ou non ($1 - T$), et pertinent soit par rapport au sous-but en cours de résolution (SBC), soit par rapport à un autre sous-but ($1 - SBC$). De plus, le modèle spécifie que soit l'une ($1P$), soit les deux ($1 - 1P$) propriétés constitutives de l'objet peuvent être représentées en mémoire. Si les deux propriétés de l'objet sont représentées ($1 - 1P$) et que cet objet est changé, alors le modèle prédit une détection correcte (DC). Par contre, si une seule des propriétés constitutives de l'objet est représentée ($1P$), elle peut soit correspondre à la propriété changée (PC) soit ne pas correspondre à la propriété changée ($1 - PC$). Nos hypothèses nous conduisent à penser que le point de vue n'aura d'effet que sur les objets directement impliqués dans le sous-but en cours de résolution. Ainsi, en ce qui concerne les autres objets de la scène (qui ne sont pas impliqués dans le sous-but en cours de résolution ou qui ne sont pas pertinents par rapport à la tâche), nous avons fixé la probabilité de représenter la propriété changée à la probabilité que donnerait le hasard, c'est-à-dire une chance sur deux ($PC = .50$). Par contre, lorsque l'objet est pertinent par rapport à la tâche et au sous-but en cours de résolution, nos hypothèses nous amènent à distinguer deux probabilités différentes de représenter la propriété changée d'un objet, selon que le changement est cohérent (PCc) ou incohérent (PCi) avec le point de vue. Dans tous les cas, si les participants ont représenté la propriété changée (PCc , PCi , ou PC), le modèle prédit une détection correcte (DC). Dans le cas où ils auraient représenté la propriété inchangée, les participants peuvent tout de même avoir la sensation qu'un changement a eu lieu sur la scène (s) ou non ($1 - s$). S'ils ont la sensation qu'un changement est survenu sur la scène, les participants peuvent deviner (g) ou non ($1 - g$) quel type de changement est survenu. Le choix du type de changement (image en plus, en moins, déplacée, ou différente) conduit à une chance sur quatre de deviner le changement qui est effectivement survenu sur l'objet, et le choix de l'icône changée (parmi les deux icônes proposées) conduit à une chance

sur deux de deviner l'icône changée : la probabilité combinatoire de détecter correctement le changement (DC) est donc estimée à une chance sur huit ($g = .125$). Alternativement, si les participants ne devinent pas le changement (*ie.*, si les participants désignent correctement le type de changement mais se trompent sur le choix de l'icône ou inversement), alors le modèle prédit une détection partielle (DP). Enfin si les participants ne sentent pas le changement ($1 - s$), ou s'ils n'ont pas représenté l'objet en mémoire ($1 - R$), le modèle prédit une non-détection (ND).

4.2. Analyse des résultats

Les réponses aux changements de l'expérience 6 ont été considérées comme des détections correctes (DC) quand les participants rapportaient correctement le type de changement (image en plus, en moins, déplacée, ou différente) *et* l'icône de l'étape changée. S'ils ne rapportaient correctement que l'un *ou* l'autre du type de changement et de l'icône modifiée, les réponses des participants étaient considérées comme des détections partielles (DP), enfin des réponses entièrement erronées ou des absences de réponse étaient considérées comme des non-détections (ND). Les nombres de détection correcte, détection partielle et non-détection, sommés pour tous les participants et tous les essais, figurent en annexe 9.

Avec $N = 792$ (132 participants * 6 essais), cinq degrés de liberté (dl), et un risque de première espèce $\alpha = .05$, un test khi-deux d'ajustement indique une valeur $X_0^2(5) = 6.61$. Cette valeur est très inférieure à la valeur critique $VC = 11.07$ donnée par la table de khi-deux pour 5 dl, ce qui indique que notre modèle, tel que l'illustre la Figure 18, avec les contraintes qui lui sont apportées ($PP = .50$; $g = .125$), est tout à fait en mesure de rendre compte des données observées dans l'expérience 6.

Le postulat à la base de notre modèle est que les représentations visuelles sont fonctionnelles et hautement transitoires, dévolues à chaque instant aux exigences de la tâche. Comme précédemment, si nous contraignons le modèle à considérer que les représentations visuelles ne relèvent ni du joug de la tâche ($T = 1 - T$), ni du joug du sous-but en cours de résolution ($SBC = 1 - SBC$), la statistique d'ajustement du modèle aux données indique respectivement $X_1^2(6) = 107.99$, et $X_2^2(6) = 10.30$. Seule la valeur X_1^2 dépasse la valeur critique pour 6 dl, $VC = 12.59$, indiquant que le modèle contraint par $T = 1 - T$ n'est pas en mesure de rendre compte des données obtenues dans l'expérience 6. La statistique d'ajustement X_2^2 reste en deçà de la valeur critique $VC(6) = 12.59$, ce qui indique que nous pouvons considérer que le modèle contraint avec $SBC = 1 - SBC$ peut expliquer nos données. Là encore, il est intéressant de pouvoir alors déterminer lequel du modèle initial ou du modèle contraint permet d'expliquer nos données de la manière la plus fine. Pour cela, on calcule un indice $\Delta X^2(\Delta dl)$ tel que $\Delta X^2 = X_2^2 - X_0^2$, assorti d'un nombre de degrés de liberté Δdl tel que $\Delta dl = dl_2 - dl_0$, qui va nous indiquer dans quelle mesure la contrainte nouvellement apportée modifie le modèle et rend compte des données. Nous obtenons $\Delta X^2(1) = 3.69$, ce qui indique, avec un risque d'erreur $\alpha = .075$, que la contrainte modifie le modèle et que le modèle initial permet de rendre compte des données plus finement que le modèle contraint.

Enfin, si nous cumulons les contraintes liées à l'immédiateté des exigences par rapport à la tâche et que nous posons $T = 1 - T$ et $SBC = 1 - SBC$, le test d'adéquation du modèle aux données indique $X_3^2(7) = 111.68$, avec 7 degrés de liberté. Cette valeur est supérieure à la valeur critique $VC = 14.07$, donnée par la table de distribution du khi-deux pour 7 dl, ce qui indique que le modèle contraint à considérer que les représentations visuelles ne subissent pas l'influence de la tâche n'est plus du tout en mesure de rendre compte des données observées.

Une des ambitions de l'expérience 6 était de permettre de rendre compte de la représentation des objets visuels lorsque les participants ne pouvaient pas tirer parti de la régularité des changements réalisés sur les scènes et recourir à une stratégie de type « ne traiter que telle ou telle propriété ». Si la probabilité de ne représenter qu'une seule des propriétés des objets de la scène estimée à 53% dans la section précédente s'avère ne pas être liée à un biais stratégique, alors on devrait retrouver ces résultats avec les données de l'expérience 6. Afin de tester cette hypothèse, nous avons contraint le présent modèle à considérer que les probabilités de représenter seulement l'une ou les deux propriétés des objets étaient équivalentes, en posant $1P = 1-1P$. Avec cette contrainte, le test de l'ajustement du modèle aux données donne une valeur $X_4^2(6) = 16.32$, avec 6 degrés de liberté. Cette valeur est supérieure à la valeur critique $VC = 12.59$, ce qui suggère que le modèle contraint à estimer qu'il y a autant de chances que les objets soient représentés en regard d'une ou des deux propriétés constitutives n'est plus en mesure de rendre compte des données de l'expérience 6.

Enfin, lorsque l'on contraint le modèle à considérer que les représentations visuelles ne répondent nullement à un critère fonctionnel et que l'on combine les contraintes $T = 1 - T$, $SBc = 1 - SBc$, et $1P = 1-1P^{108}$, la qualité de l'ajustement du modèle aux données est, avec 8 degrés de liberté, $X_5^2(8) = 128.89$. Cette valeur est bien supérieure à la valeur critique $VC = 15.51$, ce qui indique, là encore, que le modèle ainsi contraint n'est plus du tout en mesure de rendre compte des données recueillies dans l'expérience 6. Ainsi, le modèle présenté en

¹⁰⁸ En effet, notre hypothèse d'économie cognitive n'a réellement de sens qu'intégrée à un critère fonctionnel de représentation par rapport aux demandes de la tâche. Il serait en effet insensé de considérer que les ressources du système cognitif soient dilapidées dans la représentation des tous les objets de la scène qu'ils soient liés ou non à la tâche, et que le système cognitif s'évertue néanmoins à ne représenter qu'une des deux propriétés définissant ces objets.

Figure 18 semble être le plus à même d'expliquer et de rendre compte des données recueillies dans l'expérience 6.

Le Tableau 8 fournit les estimations des paramètres de notre modèle multinomial de la représentation de scènes visuelles complexes en condition de vision active (lequel est illustré en Figure 18), à partir des données de l'expérience 6.

Tableau 8. Estimations des paramètres du modèle multinomial de la représentation de scènes visuelles complexes en condition de vision active, à partir des données de l'expérience 6. R = probabilité de représenter un objet de la scène ; T = probabilité de représenter un élément pertinent par rapport à la tâche ; SBc = probabilité de représenter un élément pertinent par rapport au sous-but en cours de résolution ; 1P = probabilité de ne représenter qu'une seule des propriétés constitutives de l'objet (identité ou localisation) ; PC = probabilité de représenter la propriété changée ; PCc = probabilité de représenter la propriété changée quand le changement est cohérent avec le point de vue ; PCi = probabilité de représenter la propriété changée quand le changement est incohérent avec le point de vue ; s = probabilité de « sentir » le changement ; g = probabilité de deviner le changement ; * = maintenu constant. Déviations standard entre parenthèses. Les équations à l'origine de l'estimation des paramètres sont consultables en annexe 10

Paramètres du modèle								
R	T	SBc	1P	PCc	PCi	PC	s	g
.46	.79	.57	1.00	.64	.53	.50 *	.57	.125*
(.26)	(.03)	(.04)	(.62)	(.09)	(.09)		(72)	

Le modèle, qui montre qu'un objet de la scène a un peu moins de 50% de chances d'être représenté en mémoire, est en accord avec les théories qui postulent que le phénomène de cécité au changement résulte d'un défaut de représentation de l'information visuelle (O'Regan et al., 2000; O'Regan et al., 1999; Rensink, 2000a, 2000b, 2000d, 2001, 2002; Rensink et al., 1997, 2000; Simons, 1996; Simons & Ambinder, 2005; Simons et al., 2000; Simons & Levin, 1997; Simons & Rensink, 2005). Ensuite, le modèle montre que les objets visuels ont d'autant plus de chances d'être représentés qu'ils sont pertinents par rapport à la tâche et

pertinents par rapport au sous-but en cours. Ce résultat est, là encore, en accord avec les résultats de la littérature scientifique qui montrent que les représentations visuelles offrent des descriptions fonctionnelles des informations visuelles portées par la tâche (Hayhoe, 2000; Hayhoe et al., 2002; Hayhoe et al., 1998; Hayhoe et al., 2003; Shinoda et al., 2001; Triesch et al., 2003; Wallis & Bühlhoff, 2000). Enfin, le modèle indique que la probabilité de ne représenter que l'une des deux propriétés des objets visuels est maximale, ce qui atteste déjà un effet de la consigne sur le traitement visuel de l'information visuelle : lorsqu'une consigne individualise les propriétés des objets, le système cognitif tend à optimiser le traitement visuel de la scène en se spécialisant dans le traitement partiel des objets visuels, de manière plus appuyée qu'il ne l'aurait fait spontanément.

L'objectif principal de l'expérience 6 était de tester l'étendue du caractère fonctionnel des représentations visuelles, en étudiant l'influence de la tâche sur le traitement des objets qu'elle concerne. Pour ce faire nous avons orienté le point de vue des participants sur la tâche, au moyen d'une consigne qui met l'accent sur une des propriétés constitutives des objets visuels (leur localisation dans la scène ou leur identité), et avons observé dans quelle mesure cette accentuation, supposée orienter le point de vue des participants vers telle ou telle propriété de la scène, influait sur la détection des changements. Si, comme nous le supposons, les représentations visuelles sont construites par et pour la réalisation de la tâche, on s'attend à ce que le point de vue n'influence que le traitement visuel des objets de la scène qui sont immédiatement requis par la tâche. Par conséquent, on devrait observer que la probabilité de représenter la propriété changée évaluée par le modèle dépend du point de vue que lorsque l'objet appartient au sous-but en cours de résolution.

Un des avantages des modèles multinomiaux joints est de permettre de tester dans quelle mesure un paramètre diffère selon les conditions expérimentales. Aussi, le modèle joint

illustré en Figure 18, permet de tester l'effet de la cohérence entre le point de vue et le type de changement sur le paramètre PC (probabilité de représenter la propriété changée). Nos hypothèses nous conduisent à penser que le paramètre PC correspondant aux objets pertinents par rapport à la tâche et par rapport au sous-but courant sera influencé par la cohérence entre le point de vue et le type de changement. Un point de vue cohérent avec le type de changement devrait contribuer à augmenter la probabilité de représenter la propriété changée (PCc) alors qu'un point de vue incohérent avec le type de changement devrait contribuer à diminuer la probabilité de représenter la propriété changée (PCi). Enfin, l'on n'attend aucune influence du point de vue sur la représentation des autres objets de la scène (non pertinents par rapport au sous-but courant ou non pertinents par rapport à la tâche), aussi la probabilité de représenter la propriété changée est dans ce cas estimée par le hasard (il y a une chance sur deux que le système représente la propriété de l'objet qui sera concernée par le changement) : on pose donc $PC = .50$.

Le Tableau 8 montre que le modèle estime à 64% la probabilité d'activation du paramètre PCc, et à 53% la probabilité d'activation du paramètre PCi : ces probabilités diffèrent-elles du hasard? Lorsque l'on contraint le modèle à considérer que les paramètres PC, PCi et PCc ne diffèrent pas de la prédiction donnée par le hasard (*ie.*, lorsque l'on pose $PCc = PCi = PC = .50$), la statistique d'ajustement du modèle aux données est $X^2_6(7) = 9.75$, avec 7 dl, valeur qui est en deçà de la valeur critique donnée par la table du khi-deux pour 7 dl, $VC = 14.07$, ce qui signifie que le modèle contraint à considérer que le point de vue n'influence pas la représentation du détail de l'information visuelle est tout à fait capable d'expliquer les données observées dans l'expérience.

Cependant, dans leur revue de question théorique et empirique portant sur l'avantage présenté par l'utilisation des modèles multinomiaux dans la recherche en psychologie, Batchelder et Rieffer (1999) discutent la question de la taille de l'échantillon utilisé, trop

souvent inférieure à la taille requise pour pouvoir révéler des différences significatives entre les groupes en termes de paramètres du modèle. Un moyen de contourner le problème de la taille de l'effectif consisterait, selon eux, à estimer de façon plausible les relations complexes unissant la taille de l'échantillon, la puissance du test et le choix du risque d'erreur de première espèce (Batchelder & Riefer, 1999), de sorte à pouvoir s'affranchir du traditionnel risque d'erreur de 5%, tout en continuant à considérer que les effets observés sont significatifs. Aussi, ils stipulent que les tests d'hypothèses relatifs aux paramètres du modèle ainsi que les tests d'ajustement du modèle aux données peuvent se révéler considérablement puissants en dépit de risques d'erreur de première espèce supérieurs au conventionnel risque α de 5%, et que cette approche est particulièrement raisonnable quand le modèle s'est déjà révélé s'ajuster aux données d'études antérieures. Ils précisent qu'il existe un outil (GPOWER), proposé par Erdfelder et collaborateurs (Erdfelder, Faul, & Buchner, 1995), qui permet de calculer la puissance de différents tests statistiques, et qui peut être utilisé pour évaluer les différents niveaux de signifiante suffisants pour observer des effets significatifs en dépit d'effectifs réduits, tout en conservant un test puissant, dans les modèles MPT (Batchelder & Riefer, 1999; Smith & Bayen, 2004). Sur les recommandations de ces derniers, nous avons eu recours au programme informatique GPOWER afin d'évaluer quel risque α et quel khi-deux critique permettraient aux différences de se manifester de façon significative en fonction de nos effectifs dans les conditions changement cohérent / incohérent avec le point de vue, tout en garantissant un test d'ajustement puissant. Il en résulte que l'on peut accepter que le modèle fixé avec $PP_c = PP_i = PP = .50$ ne s'ajuste pas aux données de l'expérience tout en gardant un test puissant (puissance de test de .80) avec $\alpha = .20$ et une valeur critique $VC = 9.70$ pour 7 dl. Notre test d'ajustement donne une valeur $X^2_6(7) = 9.75$ supérieure à cette valeur critique, ce qui indique que le modèle ne s'ajuste plus aux données lorsqu'on le force à

ignorer l'influence du point de vue sur la représentation des objets visuels pertinents par rapport aux sous-but en cours de résolution.

Enfin, nous pouvons noter que le modèle estime à 57% la probabilité de « sentir » les changements et la probabilité de deviner le changement est évaluée à une chance sur huit soit 12.5%.

4.3. Discussion

Les résultats de la présente section montrent, tout d'abord, que les estimations des différents paramètres faites par le modèle à partir des résultats de l'expérience 6 restent proches des estimations faites à partir des données des expériences 4 et 5 (exception faite des paramètres 1P et s, voir plus loin), alors même que, dans l'expérience 6, la saillance perceptive inhérente à chaque changement a été contrebalancée sur les différents rangs de changement, et que plusieurs types de changements sont proposés. Ce premier résultat nous permet de rejeter définitivement une explication purement perceptive de nos résultats, et d'asseoir le rôle des buts de l'activité cognitive dans la perception de scènes visuelles complexes.

Dans cette perspective, l'objectif de cette dernière recherche était de mesurer l'étendue de l'influence de la tâche sur la représentation des objets visuels. Il s'agissait de montrer que la façon d'envisager la tâche déterminait non seulement quels objets de la scène étaient à ignorer ou à représenter en mémoire mais aussi, voire surtout, quelles propriétés des objets étaient réellement traitées et représentées en mémoire. Les résultats montrent que le point de vue des participants sur la tâche, manipulé au moyen d'une consigne, détermine littéralement les informations visuelles des objets qui seront traitées et représentées en mémoire. En effet, lorsque le type de changement susceptible d'altérer un objet de la scène n'est plus prévisible,

et que la consigne insiste sur l'une des propriétés des objets de la scène, le système cognitif se spécialise dans la représentation de cette propriété de l'objet, au détriment de l'autre propriété. Ce résultat, qui contraste avec l'estimation de la probabilité de ne représenter qu'une propriété des objets (53%) dans la section précédente¹⁰⁹, semble donc confirmer nos hypothèses : les représentations visuelles fonctionnelles fonctionnent sur un principe d'économie cognitive et tendent à ne coder que la quantité minimale d'informations requise par la tâche. Ainsi, on a pu observer que les changements qui altèrent la propriété de l'objet soulignée par la consigne (condition changement cohérent avec le point de vue, paramètre PCc) sont mieux détectés que les changements qui en altèrent l'autre propriété (condition changement incohérent avec le point de vue, paramètre PCi), ce qui montre de façon convaincante que l'orientation du point de vue sur l'une ou l'autre des deux propriétés constitutives des objets a entraîné le traitement préférentiel de cette propriété au détriment de l'autre. Aussi, lorsque le point de vue est incohérent avec le changement, on observe que la probabilité d'avoir représenté l'autre propriété que celle spécifiée par la consigne équivaut à ce que prédirait le hasard (un chance sur deux) suggérant que le point de vue invalide le traitement de la propriété qu'il ne concerne pas.

Enfin, on peut s'étonner que la probabilité correspondant à l'émergence du « visual sensing » soit si importante dans le modèle implémenté avec les données de l'expérience 6 ($s = 57\%$), comparé à la valeur que nous lui avons attribuée dans le modèle implémenté avec les données des expériences 4 et 5 ($s = 33\%$)¹¹⁰. Rappelons que, dans les travaux de Rensink (2004), la « sensation visuelle » (ou « visual sensing ») est inférée à partir du retard de

¹⁰⁹ Rappelons que pour expliquer cette probabilité de 53%, nous avons formulé une explication en terme de caractéristiques de l'expérience plutôt que du système cognitif (*ie.*, les participants apprennent et utilisent le fait qu'il n'y qu'un seul type de changement introduit dans les scènes)

¹¹⁰ Si nous contraignons le modèle à supposer que la probabilité de « visual sensing » égale 33%, alors le modèle implémenté avec les données de l'expérience 6 n'est plus ajusté aux données ($X^2(6) = 15.55$, valeur qui est supérieure à la valeur critique $VC = 12.59$)

réponses de type « perçu » sur les réponses de type « senti » chez certains participants placés dans une tâche de détection des changements. En effet, en utilisant un paradigme de type « flicker », dans lequel il intercale très rapidement les versions originale A et altérée A' de l'image, dans une séquence A, A, A', A', A, A, A', A'..., Rensink (2004) observe qu'environ un tiers des participants répond « sentir » le changement quelques secondes avant de le percevoir consciemment (on trouve des résultats similaires dans Simons, Nevarez, & Boot, sous presse). Or, le délai observé entre les réponses de type « senti » et « perçu » est interprété soit comme le reflet d'une forme de perception consciente, reposant sur des mécanismes différents de ceux qui sous-tendent le fait de « voir » (Rensink, 2000b ; 2004), soit comme une stratégie de perception basée sur un critère de détection libéral (Simons et al., sous presse). Dans notre modèle, la « sensation visuelle », fait référence au fait de « sentir » qu'un changement est survenu sur les scènes visuelles, sans pour autant l'expérimenter réellement ni être capable de l'expliquer, et elle précède le fait de deviner ou non le changement « senti ». Les proportions de « sensation visuelle » différentiellement estimées par notre modèle à partir des données des expériences 4, 5, et 6 nous semblent pouvoir résulter en partie des caractéristiques des expériences en elles-mêmes. En effet, dans les expériences 4 et 5, et à la différence de l'expérience 6, toutes les étapes des cartes étaient focalisées par l'œil avant que les participants ne commencent leur tâche. Il est possible que cette focalisation oculaire appuyée ait soit influencé le mode de traitement perceptif particulier au « sentiment visuel », soit enrichi une représentation de la scène de niveau inférieur au niveau fonctionnel (voir chapitre cinq), permettant dans les deux cas d'accroître la perception explicite du changement et de réduire la sensation du changement.

5. DISCUSSION DU CHAPITRE SEPT

Ce chapitre propose une modélisation des effets de la tâche sur la perception visuelle de scènes naturelles complexes (*ie.*, lorsque les participants doivent réaliser une tâche sur les scènes). Le modèle proposé est un modèle multinomial de la représentation de scènes visuelles complexes en condition de vision active. Les modèles multinomiaux sont des modèles statistiques simples, qui peuvent être utilisés pour mesurer, à partir de comportements observables, des processus cognitifs latents. L'utilisation de ces modèles est particulièrement adaptée à la recherche en psychologie, puisque les données recueillies résultent souvent de processus interactifs multiples, que les traitements statistiques classiques, basés sur des différences de moyennes et de variances, peinent à identifier et à déterminer. Par contraste, les modèles multinomiaux présentent l'avantageuse particularité de permettre à la fois de distinguer et de mesurer les processus cognitifs qui sous-tendent les données empiriques, et de formuler des hypothèses précises qui peuvent être testées simplement par le modèle, par le biais de tests statistiques classiques (test d'ajustement, de contingence...).

Notre modèle multinomial de la représentation de scènes visuelles complexes en condition de vision active vise à distinguer les processus cognitifs à l'origine de la perception d'une scène visuelle complexe, lorsque le traitement de cette scène est dévolu à la réalisation d'une tâche. Le modèle est implémenté à partir de plusieurs facteurs connus ou supposés influencer les représentations visuelles fonctionnelles d'une scène. Dans un premier temps, le modèle considère (i) que les objets de la scène ne sont pas forcément représentés en mémoire, même s'ils ont été préalablement focalisés par l'attention (Horowitz & Wolfe, 1998; Levin, Simons et al., 2002; O'Regan et al., 2000; Simons & Levin, 1998), (ii) que les informations visuelles ont d'autant plus de chances d'être représentées en mémoire qu'elle sont pertinentes

par rapport à la tâche (Hayhoe, 2000; Hayhoe et al., 2002; Land & Hayhoe, 2001; Most et al., 2005; Triesch et al., 2003), et immédiatement nécessaires à la résolution du sous-but en cours de traitement (Hayhoe et al., 1998). Le modèle considère, de plus, que les représentations visuelles fonctionnent sur un principe d'économie cognitive (Ballard et al., 1995), et que les représentations visuelles ne coderont que la stricte quantité d'information suffisante pour résoudre la tâche, optimisant par là même les ressources cognitives nécessaires au traitement et à la représentation en mémoire des informations. Enfin, le modèle admet la possibilité que les changements peuvent être « sentis » plutôt que perçus et qu'il existe un facteur chance de détecter correctement un changement. Enfin, il prend en compte l'éventualité qu'une même réponse puisse provenir de différents processus (par exemple, une non-détection peut résulter à la fois d'un défaut de représentation en mémoire de l'information visuelle, qui empêche systématiquement la détection des changements et d'un défaut de devinement d'un changement « senti »).

Le modèle a été implémenté avec les résultats de trois expériences mettant en jeu différents types de changements survenant à différents moments de la résolution d'une tâche de type résolution de problème. Les estimations des paramètres du modèle, qui correspondent aux différents facteurs impliqués dans la représentation en mémoire d'une information visuelle, révèlent que les représentations visuelles répondent à un critère fonctionnel. En effet, ne sont représentées que les informations visuelles non seulement requises par la tâche, mais aussi (voire surtout) immédiatement nécessaires à la résolution du sous-but dans lequel nous sommes engagés. De fait, le modèle témoigne du caractère labile du niveau fonctionnel de nos représentations visuelles, qui n'existe à chaque instant qu'en regard des exigences de la tâche. Descriptions circonstanciées de la scène, les représentations visuelles de niveau fonctionnel sont donc hautement transitoires et particulièrement dynamiques, et elles constituent un mode

de représentation privilégié des informations de la scène conformément à la tâche qui nous est imputée. De ce fait, elles garantissent le traitement optimal des informations pertinentes par rapport à la tâche. En effet, lorsque la réalisation de la tâche s'accompagne de connaissances sur les informations les plus susceptibles d'en favoriser la résolution (*ie.*, lorsqu'on manipule le point de vue des participants sur la tâche), on remarque que le système cognitif tend à ne représenter des objets visuels immédiatement pertinents que les éléments d'informations qui permettront de résoudre la tâche, au détriment des autres éléments d'informations bien qu'ils définissent tout de même ces objets visuels.

Ainsi, il convient de considérer qu'avoir à réaliser une tâche sur une scène visuelle conduit à l'émergence d'un niveau de représentation de la scène particularisé pour la résolution de la tâche. Ce niveau de représentation doit être considéré comme un instantané de la scène qui n'a de sens qu'à un moment précis de la résolution de la tâche. Une fois que l'étape de résolution deviendrait obsolète parce que résolue, les informations visuelles correspondantes déclinerait en mémoire, et seraient aussitôt remplacées par de nouvelles informations. Il serait vain alors de considérer que le niveau fonctionnel de nos représentations visuelles, qui est le niveau le plus accessible par des moyens conscients et certainement le plus susceptible de guider nos comportements, nous offre une appréhension fiable de l'ensemble de la scène, ou qu'il s'enrichirait au fur et à mesure que la tâche évolue.

TROISIÈME PARTIE :

DISCUSSION GÉNÉRALE

Discussion générale

Nous avons vu, dans le chapitre un, que les scènes visuelles naturelles peuvent être identifiées et labellisées très rapidement, dès la première fixation oculaire sur la scène (Intraub, 1981; Irwin, 1992; McCauley et al., 1980 ; Potter et al., 2002 ; Thorpe et al., 1996 ; VanRullen & Thorpe, 2001a). Néanmoins, pour que leur contenu soit représenté en mémoire et puisse, par la suite, influencer nos comportements, il est nécessaire que les différentes régions des scènes soient la cible de l'attention focalisée (Henderson & Hollingworth, 1999; Henderson et al., 2003; Intraub, 1997).

Les études sur la cécité au changement menées ces dernières années ont permis de déterminer les facteurs de la représentation des informations visuelles en mémoire : une information visuelle a d'autant plus de chances d'être focalisée par l'attention, extraite de la scène et représentée en mémoire qu'elle est pertinente pour le sujet percevant. Ainsi, les recherches menées en condition de vision passive ont révélé que les changements introduits dans les scènes sont mieux détectés lorsqu'ils altèrent des objets de la scène importants pour l'appréhender et la comprendre que lorsqu'ils altèrent d'autres objets de la scène (Mitroff & Simons, 2002; O'Regan et al., 2000; O'Regan et al., 1999; Rensink et al., 1997, 2000). Étant donné que la détection des changements est utilisée comme un indicateur des informations visuelles focalisées par l'attention et subséquemment représentées en mémoire, ces résultats indiquent que les objets d'intérêt central sont les plus à même de bénéficier de traitements attentionnel appuyés et d'être consécutivement représentés en mémoire. Par ailleurs, il a

également été mis en évidence, en condition de vision active, que la cécité au changement est modulée par la pertinence des objets changés par rapport à la tâche. En effet, les participants détectent majoritairement mieux les changements qui altèrent des objets visuels impliqués dans la tâche qui leur incombe que les changements qui altèrent d'autres objets de la scène visuelle (Hayhoe, 2000; Hayhoe et al., 1998; Shinoda et al., 2001; Wallis & Bühlhoff, 2000). Il en résulte que les objets portés par une scène visuelle ont d'autant plus de chances d'être focalisés par l'attention et représentés en mémoire qu'ils sont pertinents par rapport à la tâche.

Curieusement, la question de la représentation des scènes visuelles n'a jamais été étudiée, à notre connaissance, en regard de l'intérêt sémantique et de la pertinence du changement par rapport à la tâche. Cette question est pourtant cruciale puisque, étant donné les limites de la mémoire visuelle de travail (Irwin, 1992 ; Luck & Vogel, 1997; Melcher, 2001; Vogel et al., 2001), il est impossible que tous les objets visuels préalablement focalisés par l'œil et l'attention soient représentés en mémoire. Il est donc fondamental de pouvoir déterminer quelles informations visuelles seront extraites de la scène et suffisamment représentées en mémoire pour guider les comportements ultérieurs (Desimone & Duncan, 1995). Récemment, il a été montré que le contenu de nos représentations visuelles pouvait différer selon le type de mesure recueilli, et s'avérer soit étonnamment pauvre (Beck & Levin, 2003; Becker & Pashler, 2002; O'Regan, 2001; O'Regan et al., 2000; O'Regan et al., 1999; Rensink, 2000a, 2000b; Rensink et al., 1997, 2000), soit beaucoup plus détaillé (Hollingworth, 2004, sous presse-a, sous presse-b, sous presse-c; Hollingworth & Henderson, 2002; Hollingworth et al., 2001). Or, il est possible que ces conceptions manifestement opposées, basées sur des études qui utilisent différents paradigmes expérimentaux, évaluent, en fait, la capacité de différents niveaux de représentation de l'information visuelle. En effet, il est possible que les mesures explicites de la détection du changement, telles le report conscient du changement ou la rapidité de la réponse (utilisées dans O'Regan et al., 2000;

O'Regan et al., 1999; Rensink, 2000a, 2000b; Rensink et al., 1997, 2000, etc...) n'évaluent que le contenu des représentations visuelles qui peuvent être accédées consciemment, et que les mesures implicites de la détection de changement, telles que l'enregistrement des mouvements oculaires ou les performances à un test de choix forcé (utilisées dans Hollingworth, 2004, sous presse-a, sous presse-b, sous presse-c; Hollingworth & Henderson, 2002; Hollingworth et al., 2001), permettent de rendre compte du contenu des représentations visuelles qui sont moins accessibles par des voies conscientes (Fernandez-Duque & Thornton, 2000; Thornton & Fernandez-Duque, 2000).

L'objectif de cette thèse était donc, dans un premier temps, de déterminer dans quelle mesure l'intérêt sémantique et la pertinence de l'information visuelle par rapport à la tâche étaient codés à différents niveaux de représentation en mémoire, et dans un second temps, d'évaluer plus précisément comment le contenu des représentations visuelles de plus haut niveau évoluaient avec la résolution de la tâche.

1. SYNTHÈSE ET INTERPRÉTATION DES RÉSULTATS OBTENUS

Les expériences présentées dans le chapitre cinq avaient pour objectif d'étudier l'influence des facteurs sémantique (intérêt sémantique des objets de la scène) et fonctionnel (pertinence des objets de la scène par rapport à la tâche) sur la représentation des scènes visuelles complexes.

Dans l'ensemble, les résultats indiquent que les changements pertinents sont toujours mieux détectés que les changements non pertinents, alors que les changements centraux sont mieux détectés que les changements marginaux uniquement lorsqu'ils sont, par ailleurs, non pertinents par rapport à la tâche. Ainsi, les deux types d'informations (intérêt sémantique et

pertinence par rapport à la tâche) semblent être codés par différents niveaux de représentation : les informations liées à la compréhension de la signification de la scène seraient codées à un niveau de représentation sémantique, alors que les informations liées à la résolution de la tâche seraient codées, à un plus haut niveau, au sein d'une représentation fonctionnelle de la scène. Ce niveau de représentation fonctionnel de la scène coderait spécifiquement les informations visuelles pertinentes par rapport à la tâche, indépendamment de leur intérêt sémantique. Activé par les demandes de la tâche, il serait maintenu actif tant que la tâche ne serait pas résolue. Le niveau sémantique des représentations visuelles intégrerait, quant à lui, non seulement les informations sur les objets d'intérêt central mais également les connaissances précocement activées par la présentation de la scène, c'est-à-dire le résumé, le contexte, l'arrangement spatial et le schéma de la scène. Par ailleurs, ce niveau de représentation sémantique constituerait le niveau de référence des scènes visuelles, dont le contenu serait accessible lorsqu'il n'y aurait pas de tâche à réaliser sur les scènes ou que les informations codées au niveau fonctionnel s'avèreraient, en définitive, inappropriées à la résolution de la tâche. Enfin, on peut noter que lorsque l'on teste, par un modèle multinomial, l'hypothèse des deux niveaux de représentation avec les données de l'expérience 2, les résultats valident le modèle et renforcent ainsi l'idée que différents niveaux de représentation sous-tendent la perception de scènes visuelles complexes (cf. annexe 11).

Le chapitre six s'est, quant à lui, employé à étudier plus en détail les mécanismes du niveau fonctionnel des représentations visuelles. L'idée était qu'en observant comment sont détectés les changements qui surviennent à différents moments de la résolution d'une tâche (tâche qui nécessite que soient atteints plusieurs sous-buts avant de pouvoir atteindre le but), nous disposions d'un moyen d'évaluer la dynamique des représentations visuelles fonctionnelles. Dans ce chapitre, nous avons montré que les changements altérant les éléments visuels immédiatement nécessaires et pertinents par rapport au sous-but en cours de

résolution étaient, dans l'ensemble, mieux détectés que les changements altérant d'autres éléments de la scène, qu'ils soient devenus obsolètes parce qu'ils correspondaient à un sous-but antérieurement résolu ou, au contraire, qu'ils demeurent pertinents pour la poursuite de la tâche. Ce résultat ne pouvait ni s'expliquer par une simple exploration sélective des scènes visuelles présentées (cf. expériences 4 et 5), ni s'expliquer par un effet de la saillance perceptive des informations visuelles (cf. expérience 6). De plus, les expériences du chapitre 6 ont permis de montrer que les représentations visuelles fonctionnelles étaient construites par et pour la résolution de la tâche, et qu'elles ne codaient donc les informations visuelles qu'au fur et à mesure qu'elles étaient requises par le sous-but en cours de résolution. Ne contenant que les informations immédiatement pertinentes par rapport à la tâche le temps de leur utilisation, le niveau de représentation fonctionnel qui émerge avec la tâche constitue donc une description circonstanciée et hautement dynamique des scènes visuelles.

Dans le chapitre sept, nous avons proposé une modélisation des mécanismes sous-jacents au niveau fonctionnel des représentations visuelles, sur la base des résultats des expériences précédemment menées. Le modèle proposé est un modèle multinomial de la représentation des scènes visuelles complexes en condition de vision active, qui vise à évaluer quelles informations visuelles sont extraites, traitées et représentées par le système cognitif, et ce avec quel poids, quand notre perception des scènes est dévolue à la résolution d'une tâche complexe. Le modèle suppose que le niveau fonctionnel de nos représentations visuelles est dédié à la résolution de la tâche. Par conséquent, il suppose que nos représentations visuelles ne contiennent que l'information immédiatement et directement requise par le sous-but en cours de résolution, et, qu'ainsi, le niveau de représentation fonctionnel des scènes visuelles est hautement dynamique puisque les informations qu'il code changent au fur et à mesure que la réalisation de la tâche progresse. Enfin, le modèle considère que les représentations visuelles fonctionnelles répondent à un principe d'économie cognitive, et qu'elles tendront à

ne coder que les informations visuelles juste nécessaires pour accomplir la tâche. Ainsi, dans une perspective hautement fonctionnelle, il prédit que la quantité comme la nature des informations visuelles concernant les objets pertinents par rapport à la tâche seront toutes deux fonction du point de vue adopté pour résoudre la tâche. Enfin, il prend en considération la possibilité d'un niveau implicite de la représentation de l'information visuelle, qui se manifesterait par le biais du « visual sensing », c'est-à-dire par la capacité de certains participants à sentir qu'un changement est survenu sur la scène sans pour autant en faire l'expérience consciente.

Les données issues de trois expériences valident notre modèle. Ce faisant, elles confirment que le niveau fonctionnel de nos représentations visuelles est hautement dynamique et qu'il répond à un critère de pertinence par rapport à la tâche : les éléments de la scènes représentés en mémoire sont non seulement pertinents par rapport à la tâche mais avant tout pertinents par rapport au sous-but en cours de résolution. Notre représentation d'une scène visuelle semble donc se construire autour des informations directement et immédiatement requises par l'activité en cours et faire abstraction du reste des informations visuelles, même si ces dernières (i) restent pertinentes pour la résolution de la tâche et (ii) ont été préalablement focalisées par l'attention.

2. UN MODÈLE THÉORIQUE DES NIVEAUX DE REPRÉSENTATION DES SCÈNES VISUELLES COMPLEXES

Les résultats obtenus à l'issue de ce travail de thèse peuvent s'intégrer dans un modèle théorique plus général de la perception et la représentation des scènes visuelles complexes, en condition de vision active, condition qui nous semble constituer une approche écologique de

la perception de scènes visuelles, tant il est rare que notre exploration des scènes visuelles ne soit vouée à d'autres buts que la rêverie.

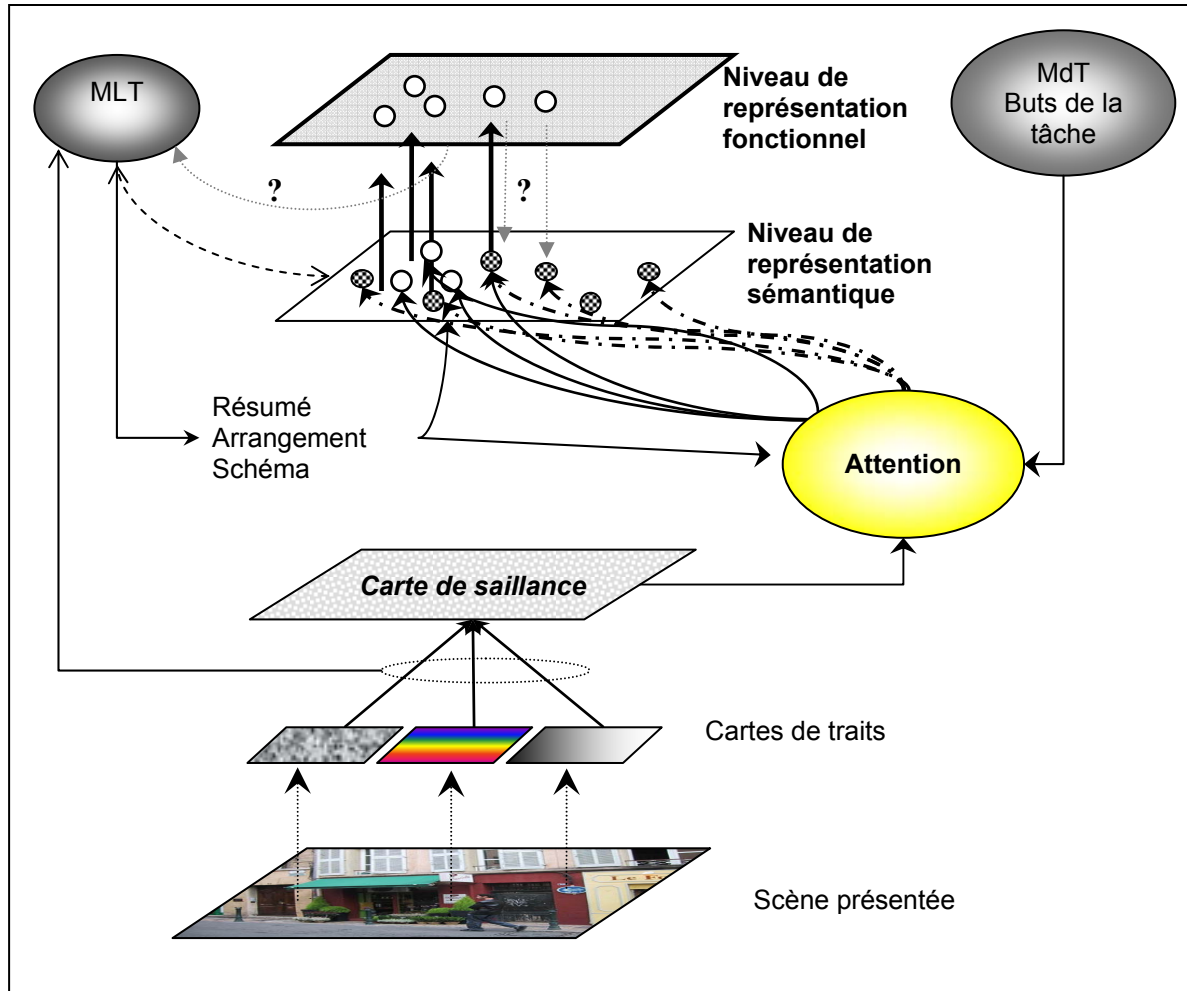


Figure 19. Architecture du modèle des niveaux de représentation d'une scène visuelle complexe

Dans ce modèle (illustré en Figure 19), les premières étapes de la perception d'une scène visuelle sont dévolues au traitement massivement parallèle des différents traits et dimensions (couleur, orientation, contraste...) basiques de la scène. L'issue de ces traitements préattentifs se matérialise par l'émergence de cartes de traits, qui codent les localisations les plus saillantes pour chacun des traits et dimensions basiques de la scène (Itti & Koch, 2000, 2001; Treisman, 1998; Treisman & Gelade, 1980; Wolfe, 1994). Ces cartes de traits sont

ensuite fusionnées en une carte de saillance telle que postulée dans les modèles computationnels du guidage de l'attention (Chauvin et al., 2002; Itti, 2005; Itti & Koch, 2000, 2001; Itti et al., 1998; Navalpakkam & Itti, 2002, 2005). La carte de saillance constitue une représentation topographique de la scène, qui code la saillance de toutes les localisations de la scène. En parallèle, l'issue des traitements préattentifs et précoces des traits et dimensions basiques entraînerait l'activation, en mémoire à long terme, de connaissances sur la scène, à savoir le résumé, le contexte, l'arrangement spatial et le schéma de la scène (Biederman, 1972; Biederman et al., 1982; Chun, sous presse; Chun & Nakayama, 2000; De Graef, 1992; Friedman, 1979; Intraub et al., 1992; Oliva, 2005). Ensemble, les connaissances sur la scène et les localisations activées dans la carte de saillance guideraient l'attention dans les scènes visuelles sur les objets d'intérêt sémantique qui permettent d'appréhender la signification de la scène (Oliva et al., 2003; Torralba, 2003). Ces objets seraient, par la suite, focalisés par l'attention, extraits et représentés en mémoire à un niveau sémantique de représentation. Ce niveau sémantique coderait non seulement le détail des objets d'intérêt central de la scène, mais serait également alimenté par les informations concernant le résumé, le contexte, l'arrangement et le schéma de la scène, ainsi que par des connaissances sur les scènes, stockées en mémoire à long terme (Hollingworth, 2004). Ce niveau de représentation sémantique émergerait relativement tôt lors du traitement visuel de la scène.

Lorsque l'on aurait à réaliser une tâche sur une scène visuelle, les buts de la tâche, stockés en mémoire à court terme se joindraient à la carte de saillance, aux connaissances sur la scène, et aux informations représentées au niveau sémantique pour guider progressivement le focus de l'attention sur les objets de la scène immédiatement pertinents par rapport à la tâche. Ces objets seraient focalisés, extraits et représentés à un niveau fonctionnel de la représentation, quel que soit leur intérêt sémantique, et ce, au fur et à mesure qu'ils seraient requis par les demandes immédiates de la tâche. Ce niveau de représentation fonctionnel

serait maintenu actif et accessible consciemment aussi longtemps que durerait la tâche. Ainsi, tant que celle-ci perdurerait, les informations visuelles les plus à même d'influencer les comportements seraient celles codées au niveau fonctionnel des représentations visuelles, c'est-à-dire les informations immédiatement pertinentes par rapport aux demandes de la tâche en cours de résolution. Une fois que le but ou le sous-but entrepris serait satisfait, l'information visuelle associée, représentée au niveau fonctionnel, déclinerait en mémoire et serait soit oubliée, soit utilisée pour enrichir la représentation sémantique et l'ensemble des connaissances sur les scènes stockées en mémoire à long terme. D'autres études comportementales nécessitent d'être entreprises afin de départager ces hypothèses.

3. PERSPECTIVES DE RECHERCHE

Nos travaux de recherche montrent que, dès les premiers regards sur une scène visuelle, les différents traitements visuels et les connaissances stockées en mémoire à long terme concourent à l'émergence d'un niveau de représentation visuelle « sémantique » de la scène, dont la principale fonction est de permettre la compréhension et l'interprétation de la scène. Lorsque l'individu aurait à réaliser une tâche sur cette scène (par exemple une tâche de recherche visuelle sur la scène), les buts de la tâche, stockés en mémoire de travail, guideraient spécifiquement l'attention visuelle sur les éléments de la scène nécessaires et pertinents pour la réalisation de la tâche en cours. Ces éléments seraient subséquemment représentés en mémoire de manière spécifique, en une représentation visuelle de niveau fonctionnel, qui se surimposerait au niveau « sémantique », sans pour autant l'annihiler. Ce niveau de représentation fonctionnel resterait actif tant que durerait la tâche.

Plusieurs aspects de ces deux niveaux de représentations visuelles, le niveau sémantique et le niveau fonctionnel, mériteraient d'être explorés plus en détail. En premier lieu, il serait intéressant de se pencher sur la chronométrie de l'émergence des niveaux sémantique et fonctionnel des représentations visuelles. En effet, le modèle théorique proposé ci-dessus postule que le niveau sémantique de la représentation visuelle émerge très précocement, et de manière automatique, dès les premiers regards sur la scène, et que les buts de la tâche entraînent ensuite l'émergence d'une représentation circonstanciée, dédiée à la résolution de la tâche. Il serait intéressant de tester la capacité de participants à détecter des changements qui surviennent sur un même objet selon qu'il est d'intérêt central ou pertinent par rapport à une tâche de recherche visuelle, et ce avec des durées de présentation extrêmement brèves. Si l'hypothèse de l'émergence anticipée du niveau sémantique sur le niveau fonctionnel s'avère justifiée, l'on devrait observer qu'il existe des durées de présentation pour lesquelles il est possible de détecter qu'un changement est survenu sur un objet d'intérêt central sans qu'il soit possible de rendre compte de changements sur des objets pertinents par rapport à la tâche. Ceci dit, ce genre d'étude nécessite que l'on s'assure que les durées utilisées sont suffisamment longues pour rendre néanmoins possible la mise en place de la tâche de recherche visuelle. L'on pourrait également proposer une tâche de recherche visuelle sur des scènes naturelles complexes, et enregistrer les patterns des mouvements oculaires. L'hypothèse de la représentation précoce du niveau de représentation sémantique prédit que l'attention devrait se poser sur les objets de la scène qui en facilitent l'appréhension avant de se focaliser sur les objets pertinents par rapport à la tâche et d'en permettre la représentation privilégiée. Cette hypothèse s'appuie, d'ailleurs, sur des résultats de Hayhoe et collaborateurs (2003), qui montrent que les premières saccades oculaires sur la scène semblent servir à échantillonner les informations présentes dans la scène, et visent à concentrer, par la suite, les fixations oculaires sur les objets pertinents par rapport à la tâche exclusivement. Or, là aussi,

il peut être délicat de parvenir à observer une orientation explicite de l'attention sur les objets d'intérêt central puisque qu'il a été montré que les objets de la scène qui sont contextuellement cohérents (et a fortiori d'intérêt central pour la compréhension de la scène?) sont les plus à même d'être traités de manière périphérique.

En second lieu, les relations entretenues par les deux types de représentations, sémantique et fonctionnelle, nécessitent d'être approfondies, notamment afin de déterminer si (et dans quelle mesure) les informations contenues dans les représentations visuelles fonctionnelles sont transférées aux représentations visuelles sémantiques, de sorte à enrichir le stock de connaissances sur les scènes visuelles stockées en mémoire à long terme, comme le suggèrent certains travaux (Hollingworth, 2004). Il a, par exemple, été montré que des participants experts détectent mieux que d'autres participants le changement d'objets d'intérêt central lorsque ces changements sont réalisés dans leur domaine d'expertise,¹¹¹ alors que rien ne distingue les performances des uns et des autres lorsque les scènes utilisées sortent du domaine d'expertise des participants¹¹² (Werner & Bjorn, 2000). Ce résultat suggère que la pertinence par rapport à la tâche des informations liées à au domaine d'expertise des participants en a certainement accentué la saillance sémantique puisque même lorsque les participants reçoivent comme consigne de regarder la scène, les informations qui sont d'ordinaires pertinentes par rapport à leur tâche sont mieux représentées en mémoire. On peut donc supposer que le niveau sémantique des représentations visuelles de ces participants s'est vu enrichi par leurs connaissances expertes. Ainsi, ce résultat constitue un premier pas dans l'acceptation de la possibilité d'une communication descendante entre le niveau fonctionnel et

¹¹¹ Dans cette expérience, les images présentées sont des scènes de football américain et les changements opérés sur les scènes sont plausibles ou non, compte tenu de la tactique de jeu. Les participants experts sont joueurs ou entraîneurs de football américain, alors que les novices ne pratiquent pas ce sport

¹¹² Dans ce cas les images sont des scènes de trafic routier

le niveau sémantique des représentations visuelles, et de l'enrichissement des connaissances sur les scènes en mémoire à long terme.

Enfin, certaines implications théoriques de notre modèle gagneraient à être appliquées aux situations à risques, et en particulier à la conduite automobile. En effet, la conduite automobile constitue, par excellence, une tâche qui conduit à l'extraction sélective de certaines informations et à l'occultation d'autres, comme en témoignent dramatiquement les accidents de type « a regardé mais n'a pas vu » (Herslund & Jorgensen, 2003), dans lesquels des éléments potentiellement dangereux n'ont été ni traités, ni représentés, ni même été en mesure d'attirer l'attention de manière exogène. Nos travaux de recherche pourraient permettre de proposer des aides à la conduite automobile, qui conduiraient, en cas de danger potentiel, les conducteurs à se désengager du niveau fonctionnel de la représentation de la scène routière. Ce désengagement permettrait aux conducteurs d'appréhender rapidement les informations potentiellement dangereuses (mais qui, étant non pertinentes à l'action de conduite qu'ils s'apprêtaient à réaliser n'étaient malheureusement pas représentées en mémoire) afin qu'ils puissent exécuter un comportement adéquat pour éviter l'accident.

Bibliographie

- Aginsky, V., & Tarr, M. J. (2000). How are different properties of a scene encoded in visual memory? *Visual Cognition*, 7, 147-162.
- Angelone, B. L., Levin, D. T., & Simons, D. J. (2003). The roles of the representation and comparison failures in change blindness. *Perception*, 32, 947-962.
- Assad, J. A. (2003). Neural coding of behavioural relevance in parietal cortex. *Current Opinion in Neurobiology*, 13, 194-197.
- Averbach, E., & Coriell, A. S. (1961). Short-term memory in vision. *The Bell System Technical Journal*, 309-328.
- Bacon-Macé, N., Macé, M. J.-M., Fabre-Thorpe, M., & Thorpe, S. (2005). The time course of visual processing: Backward masking and natural scene categorisation. *Vision Research*, 45, 1459-1469.
- Ballard, D. H., Hayhoe, M. M., & Pelz, J. (1995). Memory representations in natural tasks. *Journal of Cognitive Neuroscience*, 7, 66-80.
- Ballard, D. H., Hayhoe, M. M., Pook, P. K., & Rao, R. P. (1998). Deictic codes for the embodiment of cognition. *Behavioral and Brain Sciences*, 20, 723-767.
- Bar, M. (2004). Visual objects in context. *Nature Review Neuroscience*, 5, 617-629.
- Bastien, C. (1997). *Les connaissances de l'enfant à l'adulte*. Paris: Armand Colin.
- Batchelder, W. H., & Riefer, D. M. (1999). Theoretical and empirical review of multinomial process tree modeling. *Psychonomic Bulletin & Review*, 6, 57-86.
- Beck, M. R., & Levin, D. T. (2003). The role of representational volatility in recognizing pre- and postchange objects. *Perception and Psychophysics*, 65, 458-468.
- Becker, M. W., & Pashler, H. (2002). Volatile visual representations: failing to detect changes in recently processed information. *Psychonomic Bulletin & Review*, 9, 744-750.
- Biederman, I. (1972). Perceiving real world scene. *Science*, 177, 77-80.

- Biederman, I., Mezzanotte, R. J., & Rabinowitz, J. C. (1982). Scene perception: Detecting and judging objects undergoing relational violations. *Cognitive Psychology*, 14, 143-177.
- Blackmore, S. J., Brelstaff, G., Nelson, K., & Trocianko, T. (1995). Is the richness of our visual world an illusion? Trans-saccadic memory for complex scenes. *Perception*, 24, 1075-1081.
- Boot, W. R., Brockmole, J. R., & Simons, D. J. (sous presse). Attention capture is modulated in dual-task situations. *Psychonomic Bulletin & Review*.
- Boyce, S. J., & Pollatsek, A. (1992). Identification of objects in scenes: The role of scene background in object naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18, 531-543.
- Braun, J., & Sagi, D. (1990). Vision outside the focus of attention. *Perception and Psychophysics*, 48, 45-58.
- Bridgeman, B., van der Heijden, A. H. C., & Velichkovsky, B. M. (1994). A theory of visual stability across saccadic eye movements. *Behavioral and Brain Sciences*, 17, 247-292.
- Brockhoff, P. B., & Schlich, P. (1998). Handling replications in discrimination tests. *Food Quality and Preference*, 9, 303-312.
- Brockmole, J. R., & Henderson, J. M. (sous presse-a). Prioritization of new objects in real-world scenes: evidence from eye movements. *Journal of Experimental Psychology: Human Perception and Performance*.
- Brockmole, J. R., & Henderson, J. M. (sous presse-b). Using real-world scenes as contextual cues for search. *Visual Cognition*.
- Bruce, V., & Green, P. (1993). *La perception visuelle: physiologie, psychologie et écologie*. Grenoble: Presses Universitaires de Grenoble.
- Bullier, J. (1998). Architecture fonctionnelle du système visuel. In M. Boucart, M.-A. Hénaff & C. Belin (Eds.), *Vision: aspects perceptifs et cognitifs*. Marseille: Solal Editeurs.
- Carlson-Radvansky, L. A., & Irwin, D. E. (1995). Memory for structural information across eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 1441-1458.
- Chauvin, A., Herault, J., Marendaz, C., & Peyrin, C. (2002). Natural scene perception: visual attractors and images processing. In W. Lowe & J. Bullinasia (Eds.), *Connectionist Models of Cognition and Perception* (pp. 236-245): World Scientific Press.
- Chun, M. M. (2000). Contextual cueing of visual attention. *Trends in Cognitive Sciences*, 4, 170-178.
- Chun, M. M. (sous presse). Scene perception and memory. In D. E. Irwin & B. Ross (Eds.), *Cognitive vision*.

- Chun, M. M., & Jiang, Y. (1998). Contextual cueing: Implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology*, 36, 28-71.
- Chun, M. M., & Marois, R. (2002). The dark side of visual attention. *Current Opinion in Neurobiology*, 12, 1-6.
- Chun, M. M., & Nakayama, K. (2000). On the functional role of implicit visual memory for the adaptative deployment of attention across scenes. *Visual Cognition*, 7, 65-81.
- Chun, M. M., & Potter, M. C. (1995). A two-stage model for multiple targets detection in rapid serial visual presentation. *Journal of Experimental Psychology: Human Perception and Performance*, 21, 109-127.
- Chun, M. M., & Wolfe, J. M. (1996). Just say no: how are visual searches terminated when there is no target present? *Cognitive Psychology*, 30, 39-78.
- Chun, M. M., & Wolfe, J. M. (2001). Visual Attention. In B. Goldstein (Ed.), *Blackwell Handbook of Perception* (pp. 272-310). Oxford, UK: Blackwell Publishers Ltd.
- Colby, C. L., & Goldberg, M. E. (1999). Space and attention in parietal cortex. *Annual Review of Neuroscience*, 22, 319-349.
- Cole, G. G., Kentridge, R. W., Gellatly, A. R. H., & Heywood, C. A. (2003). Detectability of onsets versus offsets in the change detection paradigm. *Journal of Vision*, 3, 22-31.
- Corbetta, M., & Shulman, G. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Review Neuroscience*, 3, 201-215.
- De Graef, P. (1992). Scene-context effects and models of real-world perception. In K. Rayner (Ed.), *Eye movements and visual cognition: Scene perception and reading* (pp. 243-259): Springer-Verlag.
- De Graef, P. (1998). Prefixational object perception in scenes: objects popping out of schemas. In G. Underwood (Ed.), *Eye guidance in reading, driving and scene perception* (pp. 313-336). Oxford: Elsevier.
- De Graef, P., Christiaens, D., & d'Ydewalle, G. (1990). Perceptual effects of scene context on object identification. *Psychological Research*, 52, 317-329.
- De Graef, P., Lauwereyns, J., & Verfaillie, K. (2000). *Attentional orienting and scene semantics (Psychological report N°268)*: Laboratory of Experimental Psychology, University of Leuven, Belgium.
- Delorme, A., Richard, G., & Fabre-Thorpe, M. (2000). Ultra rapid categorisation of natural scenes does not rely on colour cue: a study in monkeys and humans. *Vision Research*, 40, 2187-2200.
- Dennet, D. C. (1993). *La Conscience Expliquée*. Paris: Odile Jacob.

- Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, 18, 193-222.
- Di Lollo, V. (1980). Temporal integration in visual memory. *Journal of Experimental Psychology: General*, 109, 75-97.
- Downing, P. (2000). Interaction between visual working memory and selective attention. *Psychological Science*, 11, 467-473.
- Duncan, J. (1984). Selective attention and the organisation of visual information. *Journal of Experimental Psychology: General*, 113, 501-517.
- Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review*, 96, 433-458.
- Egeth, H. E., & Yantis, S. (1997). Visual attention: control, representation, and time course. *Annual Review of Psychology*, 48, 269-297.
- Endsley, M. (1995). Toward a theory of situation awareness in dynamics systems. *Human Factors*, 37, 32-64.
- Erdfelder, E. (sous presse). *Multinomiale Modelle in der Kognitiven Psychologie: Eine Einführung [Multinomial models in cognitive psychology: An introduction]*. Cologne, Germany: Kölner Studienverlag.
- Erdfelder, E., Faul, F., & Buchner, A. (1995). GPOWER: A general power analysis program. *Behavior Research Methods, Instruments, & Computers*, 28, 1-11.
- Fabre-Thorpe, M., Delorme, A., Marlot, C., & Thorpe, S. (2001). A limit to the speed of processing in ultra rapid visual representation of novel natural scenes. *Journal of Cognitive Neuroscience*, 13, 171-180.
- Fernandez-Duque, D., & Thornton, I. M. (2000). Change detection without awareness: do explicit reports underestimate the representation of change in the visual system. *Visual Cognition*, 7, 323-344.
- Folk, C. L., Remington, R. W., & Johnston, J. C. (1992). Involuntary covert orienting is contingent on attentional control setting. *Journal of Experimental Psychology: Human Perception and Performance*, 18, 1030-1044.
- Francolini, C. M., & Egeth, H. E. (1980). On the nonautomaticity of "automatic" activation: Evidence of selective seeing. *Perception and Psychophysics*, 27, 331-342.
- Friedman, A. (1979). Framing pictures: The role of knowledge in automatized encoding and memory for gist. *Journal of Experimental Psychology: General*, 108, 316-355.
- Gilchrist, I. D., Brown, V., & Findley, J. M. (1997). Saccades without eye movements. *Nature*, 390, 130-131.

- Gottlieb, J. P., Kusunoki, M., & Goldberg, M. E. (1998). The representation of visual salience in monkey parietal cortex. *Nature*, 391, 481-484.
- Gottlieb, J. P., Kusunoki, M., & Goldberg, M. E. (2005). Simultaneous representation of saccade target and visual onset in monkey lateral intraparietal area. *Cerebral Cortex*, 15, 1198-1206.
- Grill-Spector, K., & Kanwhisher, N. (2005). Visual recognition: As soon as you know it is there, you know what it is. *Psychological Science*, 16, 152-160.
- Hayhoe, M. M. (2000). Vision using routines: A functional account of vision. *Visual Cognition*, 7, 43-64.
- Hayhoe, M. M., Ballard, D. H., Triesch, J., Shinoda, H., Aivar, P., & Sullivan, B. T. (2002). *Vision in natural and virtual environments*. Paper presented at the Eye Tracking Research and Applications Symposium, New-Orleans - Louisiana.
- Hayhoe, M. M., Bensiger, D. G., & Ballard, D. H. (1998). Task constraints in visual working memory. *Vision Research*, 38, 125-137.
- Hayhoe, M. M., Shrivastava, A., Mruczek, R., & Pelz, J. (2003). Visual memory and motor planning in a natural task. *Journal of Vision*, 3, 49-63.
- Henderson, J. M. (2003). Human gaze control during real-world scene perception. *Trends in Cognitive Sciences*, 7, 498-504.
- Henderson, J. M., & Hollingworth, A. (1999). High-level scene perception. *Annual Review of Psychology*, 50, 243-271.
- Henderson, J. M., & Hollingworth, A. (2003). Eye Movements, visual memory, and scene representation. In M. S. Peterson & G. Rhodes (Eds.), *Analytic and holistic processes in the perception of faces, objects, and scenes* (pp. 356-377). New York: JAI / Ablex.
- Henderson, J. M., Weeks, P. A., & Hollingworth, A. (1999). The effect of semantic consistency on eye movements during complex scene viewing. *Journal of Experimental Psychology: Human Perception and Performance*, 25, 210-228.
- Henderson, J. M., Williams, C. C., Castelhamo, M. S., & Falk, R. J. (2003). Eye movements and picture processing during recognition. *Perception and Psychophysics*, 65, 725-734.
- Herslund, M. B., & Jorgensen, N. O. (2003). Looked-but-failed-to-see-errors in traffic. *Accident Analysis and Prevention*, 35, 885-891.
- Hochberg, J. (1968). In the mind's eye. In R. N. Haber (Ed.), *Contemporary Theory and Research in Visual Perception* (pp. 309-331). New York: Holt, Rinehart and Winston, Inc.

- Hollingworth, A. (2003). Failure of retrieval and comparison constrain change detection in natural scenes. *Journal of Experimental Psychology: Human Perception and Performance*, 29, 388-403.
- Hollingworth, A. (2004). Constructing visual representations of natural scenes: The roles of short- and long-term visual memory. *Journal of Experimental Psychology: Human Perception and Performance*, 30, 519-537.
- Hollingworth, A. (sous presse-a). Memory for object position in natural scene. *Visual Cognition*.
- Hollingworth, A. (sous presse-b). Scene and position specificity in visual memory for objects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.
- Hollingworth, A. (sous presse-c). Visual memory for natural scenes: evidence from change blindness and visual search. *Visual Cognition*.
- Hollingworth, A., & Henderson, J. M. (2000). Semantic informativeness mediates the detection of changes in natural scenes. *Visual Cognition*, 7, 213-235.
- Hollingworth, A., & Henderson, J. M. (2002). Accurate visual memory for previously attended objects in natural scenes. *Journal of Experimental Psychology: Human Perception and Performance*, 28, 113-136.
- Hollingworth, A., Williams, C. C., & Henderson, J. M. (2001). To see and remember: visually specific information is retained in memory for previously attended objects in natural scenes. *Psychonomic Bulletin & Review*, 8, 761-768.
- Horowitz, T. S., & Wolfe, J. M. (1998). Visual search has no memory. *Nature*, 394, 575-577.
- Horowitz, T. S., & Wolfe, J. M. (2001). Search for multiple targets: Remember the targets, forget the search. *Perception and Psychophysics*, 63, 272-285.
- Hu, X. (1999). Multinomial processing tree models: an implementation. *Behavior Research Methods, Instruments, & Computers*, 31, 689-695.
- Hu, X., & Batchelder, W. H. (1994). The statistical analysis of general processing tree model with the EM algorithm. *Psychometrika*, 59, 21-47.
- Intraub, H. (1981). Rapid conceptual identification of sequentially presented pictures. *Journal of Experimental Psychology: Human Perception and Performance*, 7, 603-610.
- Intraub, H. (1997). The representation of visual scenes. *Trends in Cognitive Sciences*, 1, 217-222.
- Intraub, H. (1999). Understanding and remembering briefly glimpsed pictures: Implications for visual scanning and memory. In V. Coltheart (Ed.), *Fleeting Memories* (pp. 47-70). Cambridge, MA: The MIT Press.

- Intraub, H., Bender, R. S., & Mangels, J. A. (1992). Looking at pictures but remembering scenes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18, 180-191.
- Intraub, H., & Bodamer, J. L. (1993). Boundary extension: Fundamental aspect of pictorial representation or encoding artefact? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19, 1387-1397.
- Intraub, H., Gottesman, C. V., & Bills, A. J. (1998). Effects of perceiving and imagining scenes on memory for pictures. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24, 186-201.
- Irwin, D. E. (1992). Memory for position and identity across eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18, 307-317.
- Irwin, D. E., & Yeomans, J. M. (1986). Sensory registration and informational persistence. *Journal of Experimental Psychology: Human Perception and Performance*, 12, 343-360.
- Irwin, D. E., & Zelinsky, G. J. (2002). Eye movements and scene perception: Memory for things observed. *Perception and Psychophysics*, 64, 882-895.
- Itti, L. (2005). Models of bottom-up attention and saliency. In L. Itti, G. Rees & J. K. Tsotsos (Eds.), *Neurobiology of Attention* (pp. 576-582). San Diego, CA: Elsevier.
- Itti, L., & Arbib, M. A. (2005, June). *Visual salience facilitates entry to conscious scene representation*. Paper presented at the 9th annual meeting of the Association for the Scientific Study of Consciousness (ASSC9), Pasadena, CA.
- Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, 40, 1489-1506.
- Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nature Review Neuroscience*, 2, 194-203.
- Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20, 1254-1259.
- Jiang, Y. (2004). Time window from visual images to visual short-term memory: Consolidation or integration? *Experimental Psychology*, 51, 45-51.
- Jonides, J., & Yantis, S. (1988). Uniqueness of abrupt visual onset in capturing attention. *Perception and Psychophysics*, 43, 346-354.
- Joseph, J. S., Chun, M. M., & Nakayama, K. (1997). Attentional requirements in a "preattentive" feature search task. *Nature*, 387, 805-807.
- Kahneman, D., Treisman, A., & Gibbs, B. J. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology*, 24, 175-219.

- Karn, K. S., & Hayhoe, M. M. (2000). Memory representations guide targeting eye movements in a natural task. *Visual Cognition*, 7, 673-703.
- Kelley, T. A., Chun, M. M., & Chua, K.-P. (2003). Effects of scene inversion on change detection of targets matched for visual salience. *Journal of Vision*, 2, 1-5.
- Klein, R., Kingstone, A., & Pontefract, A. (1992). Orienting of visual attention. In K. Rayner (Ed.), *Eye movements and visual cognition: Scene perception and reading* (pp. 46-65). New York: Springer-Verlag.
- Klein, R., & Shore, D. I. (2000). Relations about modes of visual orienting. In S. Monsell & J. Driver (Eds.), *Attention and Performance XVII: Control of cognitive processes* (pp. 195-208). Cambridge: MIT Press.
- Kusunoki, M., Gottlieb, J. P., & Goldberg, M. E. (2000). The lateral intraparietal area as a salience map: the representation of abrupt onset, stimulus motion, and task-relevance. *Vision Research*, 40, 1459-1468.
- Land, M. F., & Hayhoe, M. M. (2001). In what way do eye movements contribute to every day activities? *Vision Research*, 41, 3559-3565.
- Landman, R., Spekreijse, H., & Lamme, V. A. F. (2003). Large capacity storage of integrated objects before change blindness. *Vision Research*, 43, 149-164.
- Lee, D., & Chun, M. M. (2001). What are the units of visual short-term memory: Objects or spatial locations? *Perception and Psychophysics*, 63, 253-257.
- Levin, D. T., Drivdahl, S. B., Momen, N., & Beck, M. R. (2002). False predictions about the detectability of visual changes: the role of beliefs about attention, memory, and the continuity of attended objects in causing change blindness. *Consciousness and Cognition*, 11, 507-527.
- Levin, D. T., Momen, N., & Drivdahl, S. B. (2000). Change blindness blindness: The metacognitive error of overestimating change-detection ability. *Visual Cognition*, 7, 397-412.
- Levin, D. T., & Simons, D. J. (1997). Failure to detect changes to attended objects in motion pictures. *Psychonomic Bulletin & Review*, 4, 501-506.
- Levin, D. T., Simons, D. J., Angelone, B. L., & Chabris, C. F. (2002). Memory for centrally attended changing objects in an incidental real-world change detection paradigm. *British Journal of Psychology*, 93, 289-302.
- Li, F. F., VanRullen, R., Koch, C., & Perona, P. (2002). Rapid natural scene categorisation in the near absence of attention. *PNAS*, 99, 9596-9601.
- Liversedge, S. P., & Findley, J. M. (2000). Saccadic eye movements and cognition. *Trends in Cognitive Sciences*, 4, 6-14.

- Loftus, G. R., & Mackworth, N. H. (1978). Cognitive determinants of fixation location during picture viewing. *Journal of Experimental Psychology: Human Perception and Performance*, 4, 565-572.
- Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, 390, 279-281.
- Luck, S. J., Vogel, E. K., & Shapiro, K. L. (1996). Word meanings can be accessed but not reported during the attentional blink. *Nature*, 383, 616-618.
- Ludwig, C. J. H., & Gilchrist, I. D. (2002). Stimulus-driven and goal-driven control over visual selection. *Journal of Experimental Psychology: Human Perception and Performance*, 28, 902-912.
- Mack, A., & Rock, I. (1998). Inattention blindness: Perception without attention. In R. D. Wright (Ed.), *Visual Attention* (pp. 55-75). New York Oxford: Oxford University Press.
- Mackworth, N. H., & Morandi, A. J. (1967). The gaze selects informative details within pictures. *Perception and Psychophysics*, 2, 574-552.
- Maljkovic, V., & Nakayama, K. (2000). Priming of Popout III: A short-term implicit memory system for rapid target selection. *Visual Cognition*, 571-595.
- Mandler, J. M., & Ritchey, G. H. (1977). Long-term memory for pictures. *Journal of Experimental Psychology: Human Learning and Memory*, 3, 386-396.
- Marr, D. (1982). *Vision*. New York: W. H. Freeman & Company.
- Mathis, K. (2002). Semantic inference from objects both in and out of a scene context. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28, 171-182.
- Maunsell, J. H. R. (1995). The brain's visual world: Representation of visual targets in cerebral cortex. *Science*, 270, 764-769.
- Mazer, J. A., & Gallant, J. L. (2003). Goal-related activity in V4 during free viewing visual search: evidence for a ventral stream visual salience map. *Neuron*, 40, 1241-1250.
- McCauley, C., Parmelee, C. M., Sperber, R. D., & Carr, T. H. (1980). Early extraction of meaning from pictures and its relation to conscious identification. *Journal of Experimental Psychology: Human Perception and Performance*, 6, 265-276.
- McConkie, G. W., & Currie, C. B. (1996). Visual stability across saccades while viewing complex pictures. *Journal of Experimental Psychology: Human Perception and Performance*, 22, 563-581.
- McConkie, G. W., & Zola, D. (1979). Is visual information integrated across successive fixations in reading? *Perception and Psychophysics*, 25, 221-224.

- McLeod, P., Driver, J., & Crisp, J. (1988). Visual search for conjunctions of movement and form is parallel. *Nature*, 332, 154-155.
- Melcher, D. (2001). Persistence of visual memory for scenes. *Nature*, 412, 401-402.
- Merikle, P. M., & Joordens, S. (1997). Parallels between perception without attention and perception without awareness. *Consciousness and Cognition*, 6, 219-236.
- Mitroff, S. R., & Simons, D. J. (2002). Changes are not localized before they are explicitly detected. *Visual Cognition*, 9, 937-968.
- Mitroff, S. R., Simons, D. J., & Franconeri, S. L. (2002). The siren song of implicit change detection. *Journal of Experimental Psychology: Human Perception and Performance*, 28, 798-815.
- Mitroff, S. R., Simons, D. J., & Levin, D. T. (sous presse). Nothing compares 2 views: change blindness can occur despite preserved access to the changed information. *Perception and Psychophysics*.
- Mondy, S., & Coltheart, V. (2000). Detection and identification of change in naturalistic scenes. *Visual Cognition*, 7, 281-296.
- Most, S. B., Clifford, E. R., & Simons, D. J. (2005). What you see is what you set: Sustained inattention blindness and the capture of awareness. *Psychological Review*, 112, 217-242.
- Navalpakkam, V., Arbib, M. A., & Itti, L. (2005). Attention and scene understanding. In L. Itti, G. Rees & J. K. Tsotsos (Eds.), *Neurobiology of Attention* (pp. 197-203). San Diego, CA: Elsevier.
- Navalpakkam, V., & Itti, L. (2002). A goal oriented attention guidance model. *Lecture Notes in Computer Science*, 2525, 453-451.
- Navalpakkam, V., & Itti, L. (2005). Modeling the influence of task on attention. *Vision Research*, 45, 205-231.
- Noë, A., & O'Regan, J. K. (2000). Perception, attention, and the grand illusion. *Psyche: an interdisciplinary journal of research on consciousness*, 6, retrieved from: <http://psyche.cs.monash.edu.au/v6/psyche-6-15-noe-html>.
- Noë, A., Pessoa, L., & Thompson, E. (2000). Beyond the grand illusion: What change blindness really teaches us about vision. *Visual Cognition*, 7, 93-106.
- Noles, N. S., Scholl, B. J., & Mitroff, S. R. (2005). The persistence of object files representations. *Perception and Psychophysics*, 67, 324-334.
- Oliva, A. (2005). Gist of a scene. In L. Itti, G. Rees & J. K. Tsotsos (Eds.), *Neurobiology of Attention*. San Diego, CA: Elsevier.

- Oliva, A., & Schyns, P. G. (2000). Diagnostic colors mediate scene recognition. *Cognitive Psychology*, 41, 176-210.
- Oliva, A., Torralba, A., Castelhana, M. S., & Henderson, J. M. (2003). Top-down control of visual attention in object detection. *IEEE Proceedings of the International Conference on Image Processing*, 1, 253-256.
- O'Regan, J. K. (1992). Solving the "real" mysteries of visual perception: the world as an outside memory. *Canadian Journal of Experimental Psychology*, 46, 461-488.
- O'Regan, J. K. (1994). The world as an outside iconic memory - no strong internal metric means no problem of visual stability. *Behavioral and Brain Sciences*, 17, 271-272.
- O'Regan, J. K. (2001). Thoughts on change blindness. In L. Harris & M. Jenkin (Eds.), *Vision and Attention* (pp. 281-302). New York: Springer Verlag.
- O'Regan, J. K., Deubel, H., Clark, J. J., & Rensink, R. A. (2000). Pictures change during blinks: Looking without seeing and seeing without looking. *Visual Cognition*, 7, 191-211.
- O'Regan, J. K., & Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences*, 24, 939-1011.
- O'Regan, J. K., Rensink, R. A., & Clark, J. J. (1999). Change-blindness as a result of 'mudsplashes'. *Nature*, 398, 34-35.
- Pani, J. R. (2000). Cognitive description and change blindness. *Visual Cognition*, 7, 107-126.
- Parkhurst, D., Law, K., & Niebur, E. (2002). Modeling the role of salience in the allocation of overt visual attention. *Vision Research*, 42, 107-123.
- Parkhurst, D., & Niebur, E. (2004). *Salience and overt visual attention*. Paper presented at the Early Cognitive Vision Workshop, Scotland.
- Pashler, H. (1988). Familiarity and visual change detection. *Perception and Psychophysics*, 44, 369-378.
- Pashler, H. (1998). *Attention*. Philadelphia: Taylor & Francis Press.
- Pelz, J., Hayhoe, M. M., & Loeber, R. (2001). The coordination of eye, head, and hand movements in a natural task. *Experimental Brain Research*, 139, 266-277.
- Peterson, M. S., & Kramer, A. F. (2001). Attentional guidance of the eyes by contextual information and abrupt onset. *Perception and Psychophysics*, 63, 1239-1249.
- Phillips, W. A. (1974). On the distinction between sensory storage and short-term visual memory. *Perception and Psychophysics*, 16, 283-290.
- Poitrenaud, S. (1995). The PROCOPE semantic network: an alternative to action grammars. *International Journal of Human-Computer Studies*, 42, 31-69.

- Posner, M. I. (1980). Orientating of attention. *Quarterly Journal of Experimental Psychology*, 32A, 3-25.
- Potter, M. C. (1976). Short-term conceptual memory for pictures. *Journal of Experimental Psychology: Human Learning and Memory*, 2, 509-522.
- Potter, M. C. (1993). Very short-term conceptual memory. *Memory & Cognition*, 21, 156-161.
- Potter, M. C. (1999). Understanding sentences and scenes: The role of conceptual short-term memory. In V. Coltheart (Ed.), *Fleeting Memories* (pp. 12-46). Cambridge, MA: The MIT Press.
- Potter, M. C., Staub, A., & O'Connor, D. H. (2004). Pictorial and conceptual representation of glimpsed pictures. *Journal of Experimental Psychology: Human Perception and Performance*, 30, 478-489.
- Potter, M. C., Staub, A., Rado, J., & O'Connor, D. H. (2002). Recognition memory for briefly presented pictures: The time course of rapid forgetting. *Journal of Experimental Psychology: Human Perception and Performance*, 28, 1163-1175.
- Prinzmetal, W., Presti, D. E., & Posner, M. I. (1986). Does attention affect visual feature integration? *Journal of Experimental Psychology: Human Perception and Performance*, 12, 361-369.
- Pylyshyn, Z. (1999). Is vision continuous with cognition? The case for cognitive impenetrability of visual perception. *Behavioral and Brain Sciences*, 22, 341-423.
- Rao, R. P., Zelinsky, G. J., Hayhoe, M. M., & Ballard, D. H. (2002). Eye movements in iconic visual search. *Vision Research*, 42, 1447-1463.
- Rensink, R. A. (2000a). The dynamic representation of scenes. *Visual Cognition*, 7, 17-42.
- Rensink, R. A. (2000b). Seeing, sensing, and scrutinizing. *Vision Research*, 40, 1469-1487.
- Rensink, R. A. (2000c). Visual search for changes: a probe into the nature of attentional processing. *Visual Cognition*, 7, 345-376.
- Rensink, R. A. (2000d). When good observers go bad. *Psyche: an interdisciplinary journal of research on consciousness*, 6, retrieved from: <http://psyche.cs.monash.edu.au/v6/psyche-6-09-rensink.html>.
- Rensink, R. A. (2001). Change Blindness: Implications for visual attention. In M. Jenkin & L. Harris (Eds.), *Vision & Attention* (pp. 169-188). New-York: Springer.
- Rensink, R. A. (2002). Change detection. *Annual Review of Psychology*, 53, 245-277.
- Rensink, R. A. (2004). Visual sensing without seeing. *Psychological Science*, 15, 27-32.

- Rensink, R. A., O'Regan, J. K., & Clark, J. J. (1997). To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science*, 5, 368-375.
- Rensink, R. A., O'Regan, J. K., & Clark, J. J. (2000). On the failure to detect changes in scenes across brief interruptions. *Visual Cognition*, 7, 127-145.
- Repin, I. (Artist). (1884-1888). *The unexpected visitor*
- Reynolds, J. H., Chelazzi, L., & Desimone, R. (1999). Competitive mechanisms subserve attention in Macaque Areas V2 and V4. *The Journal of Neuroscience*, 19, 1736-1753.
- Richard, J.-F. (1999). Comportements, Buts et représentations. *Psychologie Française*, 44, 75-90.
- Sanocki, T. (2003). Representation and perception of scenic layout. *Cognitive Psychology*, 47, 43-86.
- Schlich, P. (1993). Risks table for discrimination tests. *Food Quality and Preference*, 4, 141-151.
- Schlich, P., Dacremont, C., & Brockhoff, P. B. (2000). Application of replicated difference testing. *Food Quality and Preference*, 11, 43-46.
- Scholl, B. J. (2000). Attenuated change blindness for exogenously attended items in a flicker paradigm. *Visual Cognition*, 7, 377-396.
- Scholl, B. J. (2001). Objects and attention: the state of the art. *Cognition*, 80, 1-4.
- Schyns, P. G., & Oliva, A. (1994). From blobs to boundary edges: evidence for time- and spatial-scale-dependent scene recognition. *Psychological Science*, 5, 195-200.
- Scott-Brown, K. C., Baker, M. R., & Orbach, H. S. (2000). Comparison blindness. *Visual Cognition*, 7, 253-267.
- Shapiro, K. L. (2000). Change blindness: theory or paradigm? *Visual Cognition*, 7, 83-91.
- Shinoda, H., Hayhoe, M. M., & Shrivastava, A. (2001). What controls attention in natural environments. *Vision Research*, 41, 3535-3546.
- Shore, D. I., & Klein, R. (2000). The effects of scene inversion on change blindness. *The Journal of General Psychology*, 127, 27-43.
- Simons, D. J. (1996). In sight, out of mind: when object representation fail. *Psychological Science*, 7, 301-305.
- Simons, D. J. (2000a). Attentional capture and inattention blindness. *Trends in Cognitive Sciences*, 4, 147-155.
- Simons, D. J. (2000b). Current approaches of change blindness. *Visual Cognition*, 7, 1-15.

- Simons, D. J., & Ambinder, M. S. (2005). Change blindness: theory and consequences. *Current Directions in Psychological Science*, 14, 44-48.
- Simons, D. J., & Chabris, C. F. (1999). Gorillas in our midst: sustained inattention blindness for dynamic events. *Perception*, 28, 1059-1074.
- Simons, D. J., Chabris, C. F., Schnur, T., & Levin, D. T. (2002). Evidence for preserved representations in change blindness. *Consciousness and Cognition*, 11, 78-97.
- Simons, D. J., Franconeri, S. L., & Reimer, R. L. (2000). Change blindness in the absence of a visual disruption. *Perception*, 29, 1143-1154.
- Simons, D. J., & Levin, D. T. (1997). Change blindness. *Trends in Cognitive Sciences*, 1, 261-267.
- Simons, D. J., & Levin, D. T. (1998). Failure to detect changes to people during real-world interaction. *Psychonomic Bulletin & Review*, 5, 644-649.
- Simons, D. J., Mitroff, S. R., & Franconeri, S. L. (2003). Scene perception: what we can learn from visual integration and change detection. In M. S. Peterson & G. Rhodes (Eds.), *Perception of Faces, Objects, and Scenes: Analytic and Holistic Processes* (pp. 335-355). New York Oxford: Oxford University Press.
- Simons, D. J., Nevarez, G., & Boot, W. R. (sous presse). Visual sensing is seeing: why "mindsight", in hindsight, is blind. *Psychological Science*.
- Simons, D. J., & Rensink, R. A. (2005). Change blindness: past, present, and future. *Trends in Cognitive Sciences*, 9, 16-20.
- Smilek, D., Eastwood, J. D., & Merikle, P. M. (2000). Does unattended information facilitate change detection? *Journal of Experimental Psychology: Human Perception and Performance*, 26, 480-487.
- Smith, R. E., & Bayen, U. J. (2004). A multinomial model of event-based prospective memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30, 756-777.
- Snyder, L. H., Batista, A. P., & Andersen, R. A. (1997). Coding of intention in the posterior parietal cortex. *Nature*, 386, 167-170.
- Spitzer, H., Desimone, R., & Moran, J. (1988). Increased attention enhances both behavioural and neuronal performance. *Science*, 240, 338-340.
- Standing, L. (1973). Learning 10,000 pictures. *Quarterly Journal of Experimental Psychology*, 25A, 207-222.
- Stolz, J. A., & Jolicoeur, P. (2004). Changing features do not guide attention in change detection: evidence from a spatial cueing paradigm. *Psychonomic Bulletin & Review*, 11, 870-875.

- Tatler, B. W., Gilchrist, I. D., & Rusted, J. (2003). The time course of abstract visual representation. *Perception*, 32, 579-592.
- Theeuwes, J. (1994). Stimulus-driven capture and attentional set: Selective search for color and visual abrupt onsets. *Journal of Experimental Psychology: Human Perception and Performance*, 20, 799-806.
- Theeuwes, J. (2005). Irrelevant singleton capture attention. In L. Itti, G. Rees & J. K. Tsotsos (Eds.), *Neurobiology of Attention* (pp. 418-424). San Diego, CA: Elsevier.
- Theeuwes, J., Kramer, A. F., Hahn, S., & Irwin, D. E. (1998). Our eyes do not always go where we want them to go: Capture of the eyes by new objects. *Psychological Science*, 9, 379-385.
- Theeuwes, J., Kramer, A. F., & Kingstone, A. (2004). Attentional capture modulates perceptual sensitivity. *Psychonomic Bulletin & Review*, 11, 551-554.
- Thornton, I. M., & Fernandez-Duque, D. (2000). An implicit measure of undetected change. *Spatial Vision*, 14, 21-44.
- Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, 381, 520-522.
- Tijus, C. A., & Reeves, A. (2004). Rapid iconic erasure without masking. *Spatial Vision*, 17, 483-495.
- Torralba, A. (2003). Modeling global scene factors in attention. *Journal of Optical Society of America*, 20, 1407-1418.
- Torralba, A., & Oliva, A. (2003). Statistics of natural images categories. *Computation in Neural Systems*, 14, 391-412.
- Treisman, A. (1998). The perception of features and objects. In R. D. Wright (Ed.), *Visual Attention* (pp. 26-54). New York Oxford: Oxford University Press.
- Treisman, A., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, 12, 97-136.
- Treue, S. (2003). Visual attention: the where, what, how and why of saliency. *Current Opinion in Neurobiology*, 13, 428-432.
- Triesch, J., Ballard, D. H., Hayhoe, M. M., & Sullivan, B. T. (2003). What you see is what you need. *Journal of Vision*, 3, 86-94.
- Tversky, B., & Hemenway, K. (1983). Categories of environmental scenes. *Cognitive Psychology*, 15, 121-149.
- Tversky, B., & Hemenway, K. (1984). Objects, parts, and categories. *Journal of Experimental Psychology: General*, 113, 169-193.

- Ungerleider, L. G., & Haxby, J. (1994). What and where in the human brain. *Current Opinion in Neurobiology*, 4, 157-165.
- van Zoest, W., Donk, M., & Theeuwes, J. (2004). The role of stimulus-driven and goal-driven control in saccadic visual selection. *Journal of Experimental Psychology: Human Perception and Performance*, 30, 746-759.
- VanRullen, R., & Koch, C. (2003). Competition and selection during visual processing of natural scenes and objects. *Journal of Vision*, 3, 75-85.
- VanRullen, R., & Thorpe, S. (2001a). Is it a bird? Is it a plane? Ultra-rapid categorisation of natural and artificial objects. *Perception*, 30, 655-668.
- VanRullen, R., & Thorpe, S. (2001b). The time course of visual processing: from early perception to decision-making. *Journal of Cognitive Neuroscience*, 13, 454-461.
- Vogel, E. K., Woodman, G. F., & Luck, S. J. (2001). Storage of features, conjunctions and objects in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance*, 27, 92-114.
- Wallis, G., & Bühlhoff, H. H. (2000). What's scene and not seen: influences of movement and task upon what we see. *Visual Cognition*, 7, 175-190.
- Ward, R., & Duncan, J. (1996). The time-course of visual attention. *Cognitive Psychology*, 30, 79-109.
- Watanabee, K. (2003). Differential effect of distractor timing on localizing versus identifying visual changes. *Cognition*, 88, 243-257.
- Werner, S., & Bjorn, T. (2000). Is "change blindness" attenuated by domain-specific expertise? An expert-novice comparison of change detection in football images. *Visual Cognition*, 7, 163-173.
- Wilken, P., & Ma, W. J. (2004). A detection theory account of change detection. *Journal of Vision*, 4, 1120-1135.
- Williams, P., & Simons, D. J. (2000). Detecting changes in novel, complex, three-dimensional objects. *Visual Cognition*, 7, 297-322.
- Wolfe, J. M. (1994). Guided search 2.0: A revised model of visual search. *Psychonomic Bulletin & Review*, 1, 202-238.
- Wolfe, J. M. (1998). What do you know about what you saw? *Current Biology*, 8, R303-R304.
- Wolfe, J. M. (1999). Inattentional amnesia. In V. Coltheart (Ed.), *Fleeting Memories* (pp. 71-94). Cambridge, MA: MIT Press.
- Wolfe, J. M., & Bennett, S. C. (1997). Preattentive object files: Shapeless bundles of basic features. *Vision Research*, 37, 25-43.

- Wolfe, J. M., Cave, K. R., & Franzel, S. L. (1989). Guided search: an alternative to the feature integration model for visual search. *Journal of Experimental Psychology: Human Perception and Performance*, 15, 419-433.
- Wolfe, J. M., & Horowitz, T. S. (2004). What attributes guide the deployment of visual attention and how do they do it? *Nature Review Neuroscience*, 5, 1-7.
- Wright, R. D. (1998). *Visual Attention*. New York: Oxford University Press.
- Yantis, S. (2003). To see is to attend. *Science*, 299, 54-56.
- Yantis, S. (2005). How visual salience wins the battle for awareness. *Nature Neuroscience*, 8, 975-977.
- Yantis, S., & Egeth, H. E. (1999). On the distinction between visual salience and stimulus-driven attentional capture. *Journal of Experimental Psychology: Human Perception and Performance*, 25, 661-676.
- Yantis, S., & Johnston, J. C. (1990). On the locus of visual selection: Evidence from focused attention tasks. *Journal of Experimental Psychology: Human Perception and Performance*, 16, 135-149.
- Yantis, S., & Jonides, J. (1990). Abrupt visual onsets and selective attention: voluntary versus automatic allocation. *Journal of Experimental Psychology: Human Perception and Performance*, 16, 121-134.
- Yarbus, A. L. (1967). *Eye Movements and Vision*. New-York: Plenum Press.
- Zelinsky, G. J., Rao, R. P., Hayhoe, M. M., & Ballard, D. H. (1997). Eye movements reveal the spatiotemporal dynamics of visual search. *Psychological Science*, 8, 448-453.

Annexes

ANNEXE 1 : Construction du matériel de l'expérience 1

ANNEXE 2 : Exemple de chaque combinaison de changement (expérience 2)

ANNEXE 3a : Cartes utilisées dans les expériences 3, 4, 5, et 6

ANNEXE 3b : Masque utilisée (expériences 3, 4, 5, 6)

ANNEXE 4 : Exemple de changement global, dans la condition « disparition » (expérience 3).

ANNEXE 5 : Nombre de réponses en fonction du rang du changement et des expériences
implémentées dans le modèle MPT (expériences 4 et 5)

ANNEXE 6 : Équations générées par le modèle MPT (expériences 4 et 5)

ANNEXE 7 : Arbres des tâches utilisés dans l'expérience 6

ANNEXE 8 : Plan en carré latin de l'expérience 6

ANNEXE 9 : Nombre de réponses en fonction du rang du changement et des expériences
implémentées dans le modèle MPT (expérience 6)

ANNEXE 10 : Équations générées par le modèle MPT (expérience 6)

ANNEXE 11 : Un modèle multinomial des niveaux de représentation des scènes visuelles
complexes

ANNEXE 1 : CONSTRUCTION DU MATÉRIEL DE L'EXPÉRIENCE 1

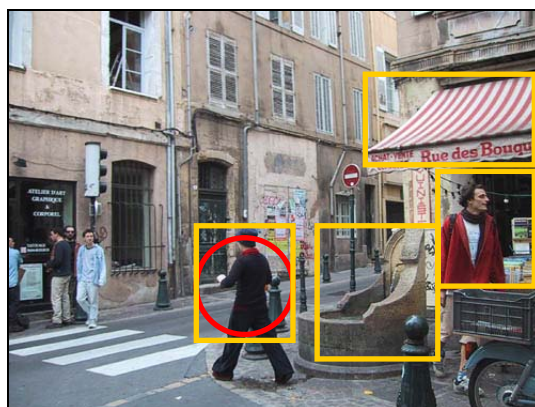
Répartition des changements en fonction de leur pertinence par rapport à la tâche et de leur intérêt sémantique. C = central ; M = marginal ; P = pertinent ; NP = non pertinent

	Intérêt sémantique du changement	
	C	M
Pertinence du changement		
P	1 disparition 2 additions	2 disparitions 2 additions
NP	2 disparitions 2 additions	1 disparition 1 addition 1 changement de couleur

ANNEXE 2 : EXEMPLE DE CHAQUE COMBINAISON DE CHANGEMENT

(EXPÉRIENCE 2)

Les cercles montrent l'objet changé, les rectangles désignent les objets pertinents par rapport à la tâche (colonne de gauche), et les objets d'intérêt central (colonne de droite).



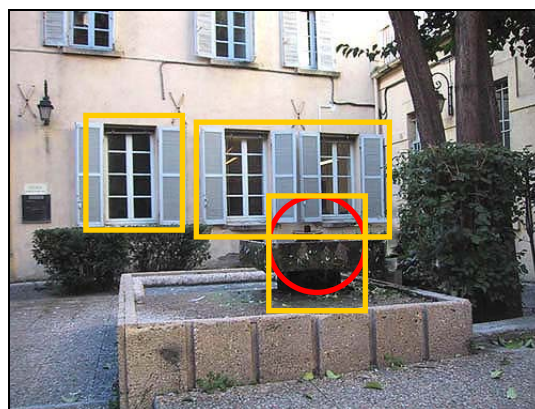
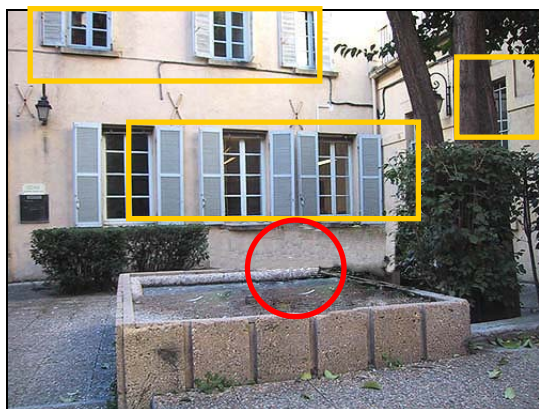
Changement pertinent central (changement de couleur)

Consigne : Cliquer sur les êtres humains



Changement pertinent marginal (disparition)

Consigne : Cliquer sur les plots



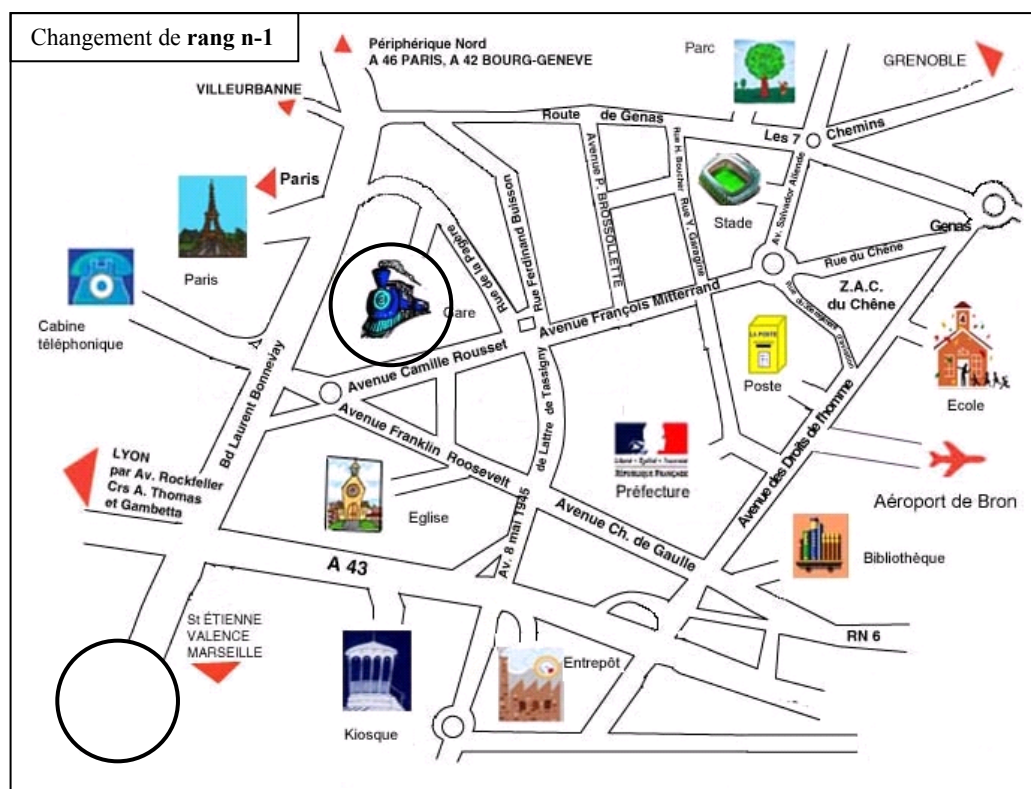
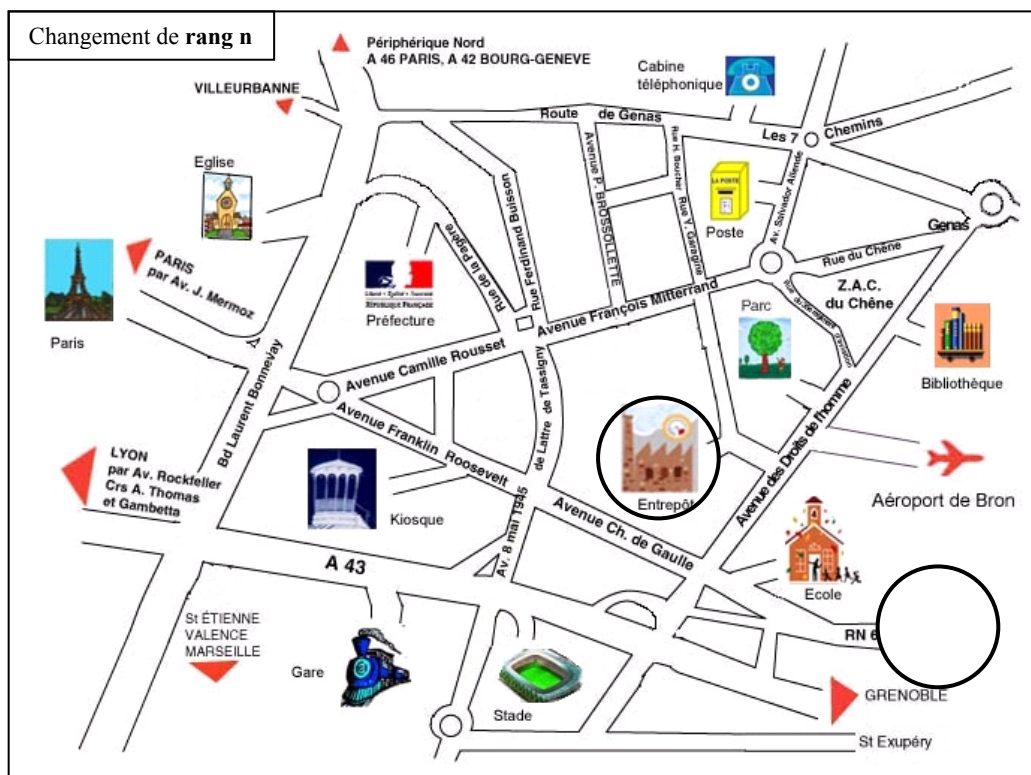
Changement non pertinent central (addition)
Consigne : Cliquer sur les fenêtres



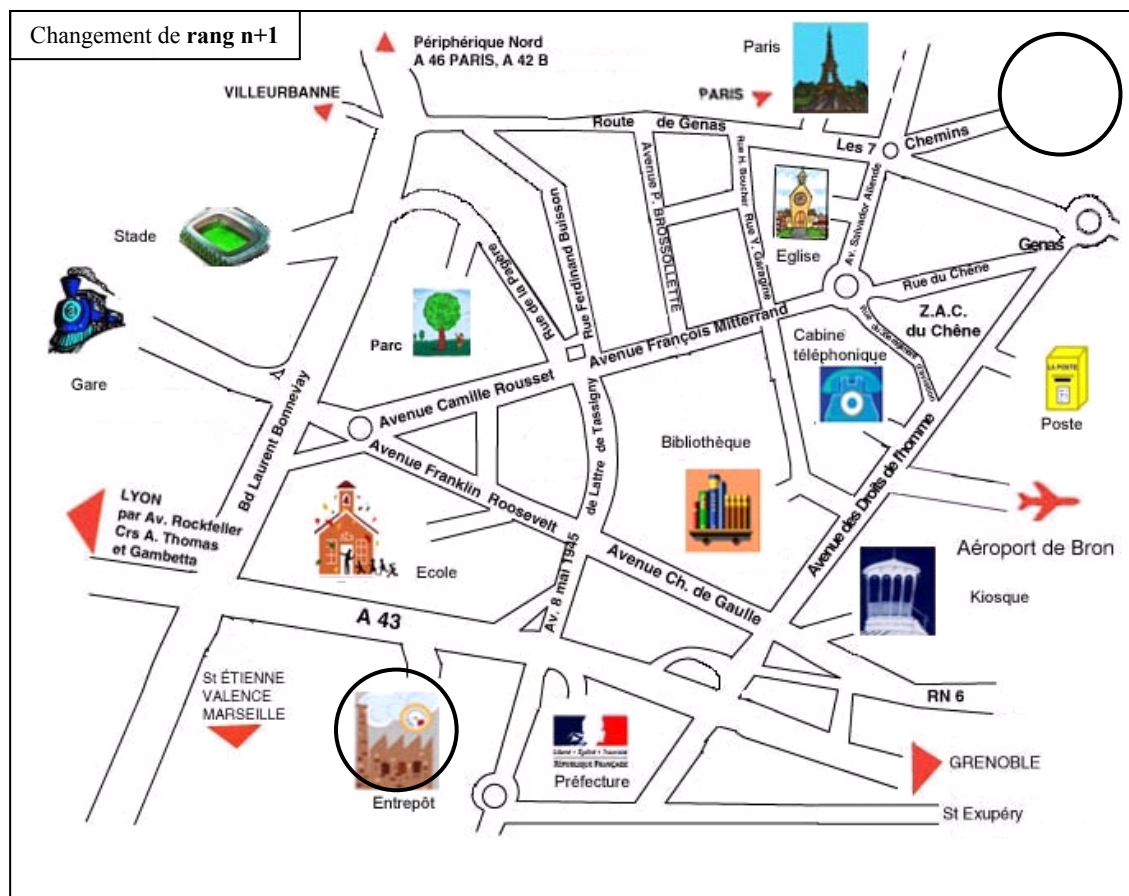
Changement non pertinent marginal (déplacement)
Consigne : Cliquer sur les enseignes

ANNEXE 3A : CARTES UTILISÉES (EXPÉRIENCES 3, 4, 5, 6)

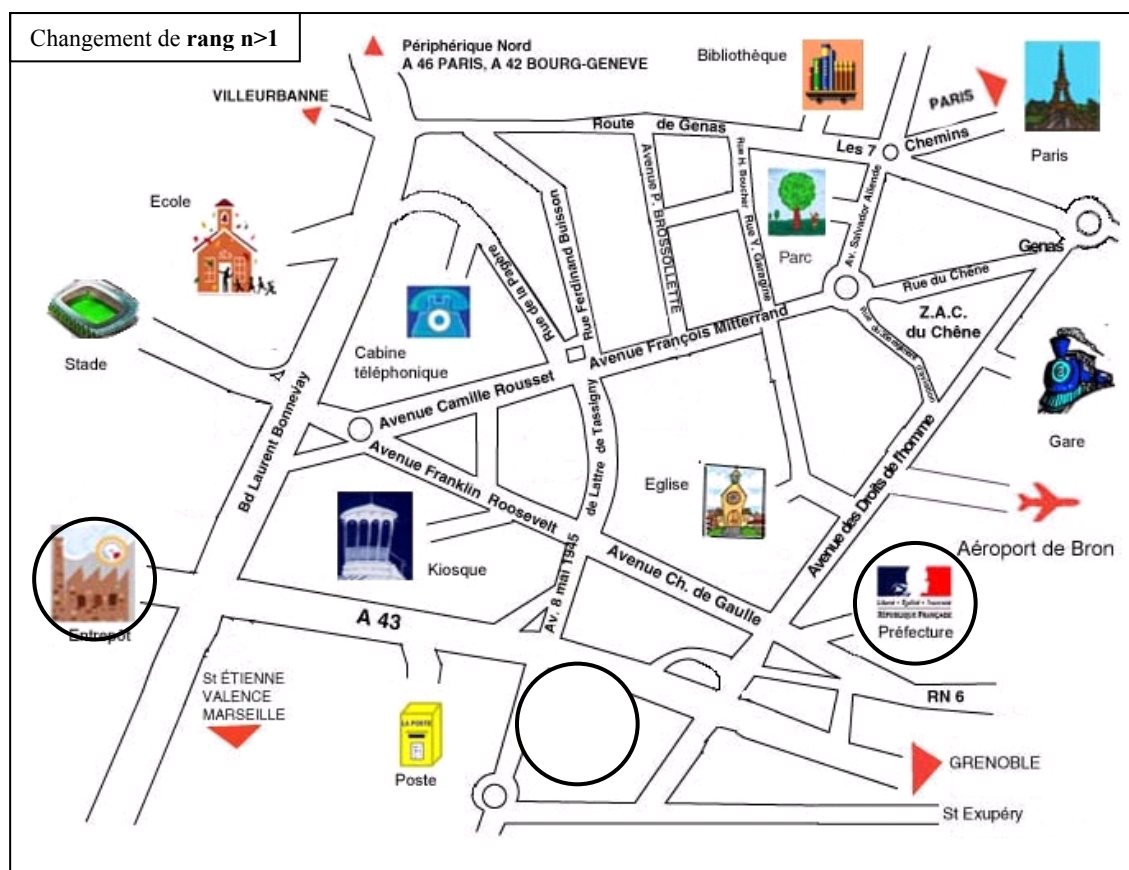
Les cercles montrent les changements effectués. Dans le cas d'un déplacement, l'étape encerclée migre vers le cercle vide, dans le cas d'une disparition, elle est effacée de la carte.

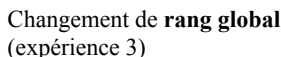
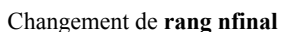


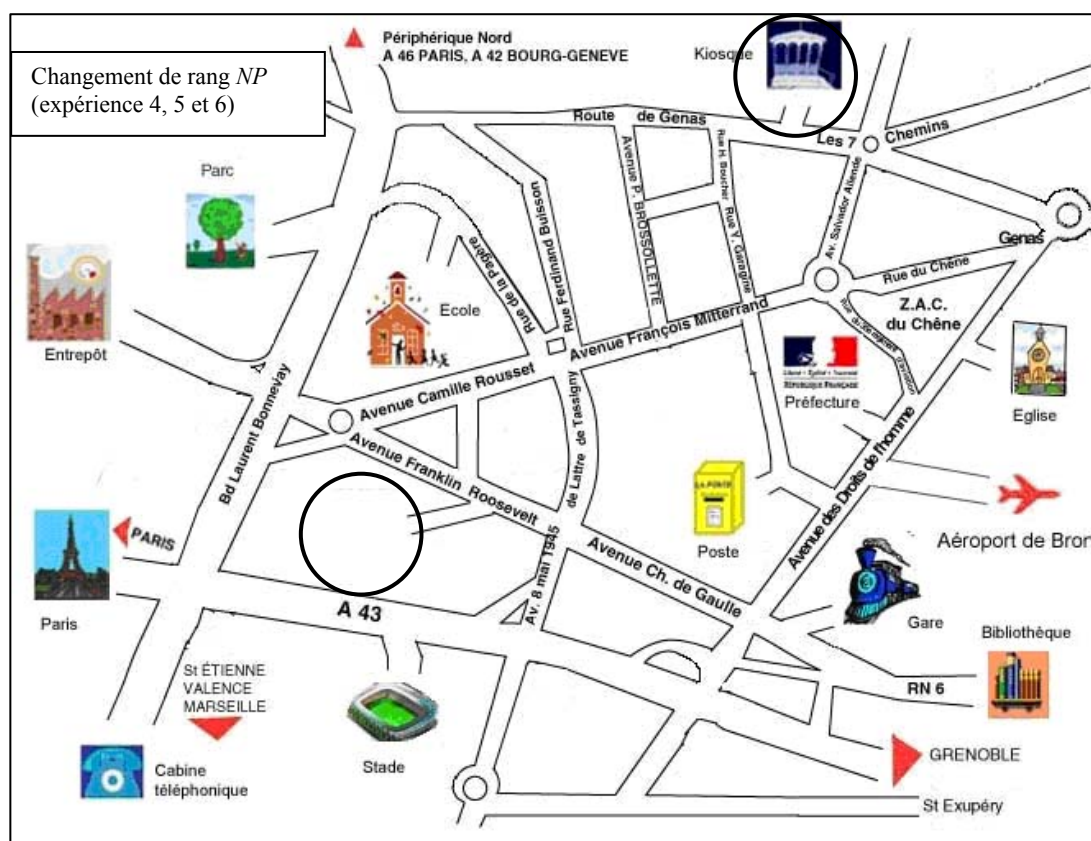
Changement de rang n+1



Changement de rang n+1







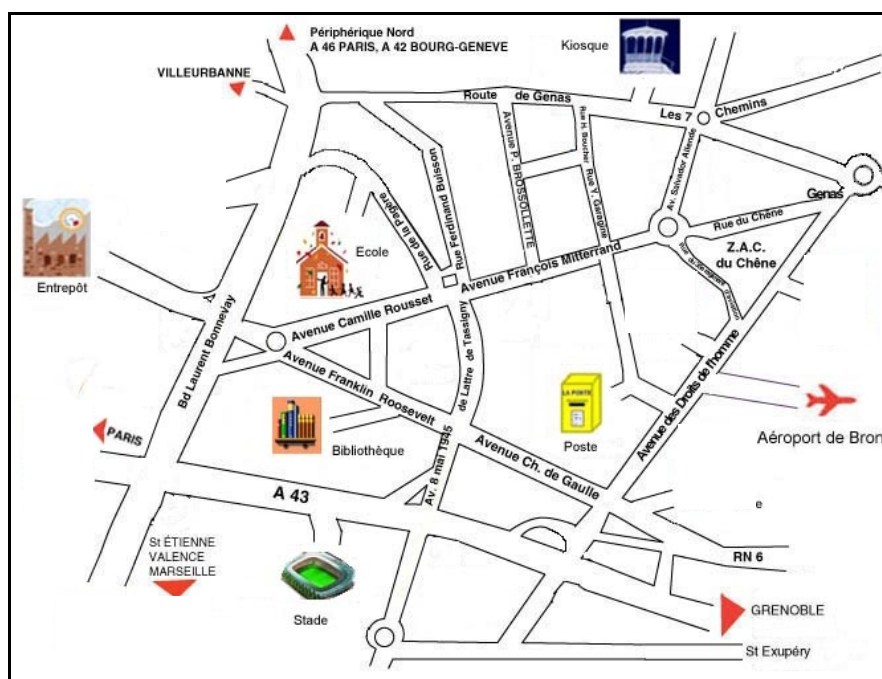
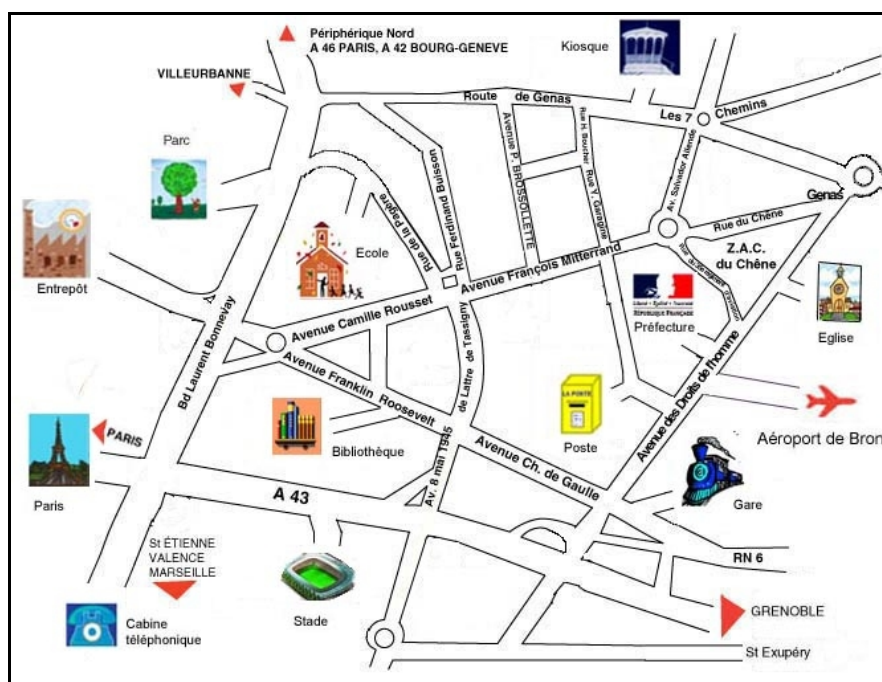
ANNEXE 3B : MASQUE UTILISÉE (EXPÉRIENCES 3, 4, 5, 6)



ANNEXE 4 : EXEMPLE DE CHANGEMENT GLOBAL, DANS LA CONDITION

« DISPARITION » (EXPÉRIENCE 3).

En haut la carte avant disparition des étapes ; en bas, la même carte après disparitions des six étapes, dans la condition « 6 changements simultanés »



**ANNEXE 5 : NOMBRE DE RÉPONSES EN FONCTION DU RANG DU CHANGEMENT ET
DES EXPÉRIENCES IMPLÉMENTÉ DANS LE MODÈLE MPT (EXPÉRIENCES 4 & 5)**

Pertinence du changement par rapport à la tâche				
		Oui (T)	Non (1 – T)	
Pertinence par rapport au sous-but en cours				
		Oui (SBc)	Non (1 – SBc)	total
Expérience 4				
DC	20	12	11	43
DP	1	1	1	3
ND	27	35	36	98
Expérience 5				
DC	22	8	9	39
DP	1	1	2	4
ND	25	39	37	101

Note : Étant donné que nous ne pouvons pas évaluer quelle est la part des non-détections (ND₁, ND₂, et ND₃) recueillies qui est due à un défaut de représentation de l'information (1 – R) ou à un défaut de « visual sensing » (1 – s), nous avons implémenté dans le modèle le nombre total de non-détections, et laissons au modèle le soin d'évaluer la part des non-détections qui découlent d'un défaut de représentation et d'un défaut de « visual sensing ».

ANNEXE 6 : ÉQUATIONS GÉNÉRÉES PAR LE MODÈLE MPT (EXPÉRIENCES 4 & 5)

Ces équations définissent la probabilité d'une catégorie de réponse (DC = détection correcte ; DP = détection partielle ; ND = non-détection) en fonction du type d'item (T = pertinent par rapport à la tâche ; 1 - T = non pertinent par rapport à la tâche ; SBc = pertinent par rapport au sous-but en cours de résolution ; 1 - SBc = non pertinent par rapport au sous-but en cours de résolution). Les données observées pour chaque catégorie de réponses figurent dans le tableau de l'annexe 5.

Expérience 4

$$\text{Pr (DC / T, SBc)} = R4 * T * SBc * (1P) * PC + R4 * T * SBc * (1P) * (1-SBc) * s * g + R4 * T * SBc * (1-1P)$$

$$\text{Pr (DC / T, 1 - SBc)} = R4 * T * (1-SBc) * (1P) * PC + R4 * T * (1-SBc) * (1P) * (1-SBc) * s * g + R4 * T * (1-SBc) * (1-1P)$$

$$\text{Pr (DC / 1 - T)} = R4 * (1-T) * (1P) * PC + R4 * (1-T) * (1P) * (1-SBc) * s * g + R4 * (1-T) * (1-1P)$$

$$\text{Pr (DP / T, SBc)} = R4 * T * SBc * (1P) * (1-SBc) * s * (1-g)$$

$$\text{Pr (DP / T, 1 - SBc)} = R4 * T * (1-SBc) * (1P) * (1-SBc) * s * (1-g)$$

$$\text{Pr (DP / 1 - T)} = R4 * (1-T) * (1P) * (1-SBc) * s * (1-g)$$

$$\text{Pr (ND)} = R4 * T * SBc * (1P) * (1-SBc) * (1-s) + R4 * T * (1-SBc) * (1P) * (1-SBc) * (1-s) + R4 * (1-T) * (1P) * (1-SBc) * (1-s) + (1-R4)$$

Expérience 5

$$\Pr (DC / T, SBc) = R5 * T * SBc * (1P) * PC + R5 * T * SBc * (1P) * (1-SBc) * s * g + R5 * T * SBc * (1-1P)$$

$$\Pr (DC / T, 1 - SBc) = R5 * T * (1-SBc) * (1P) * PC + R5 * T * (1-SBc) * (1P) * (1-SBc) * s * g + R5 * T * (1-SBc) * (1-1P)$$

$$\Pr (DC / 1 - T) = R5 * (1-T) * (1P) * PC + R5 * (1-T) * (1P) * (1-SBc) * s * g + R5 * (1-T) * (1-1P)$$

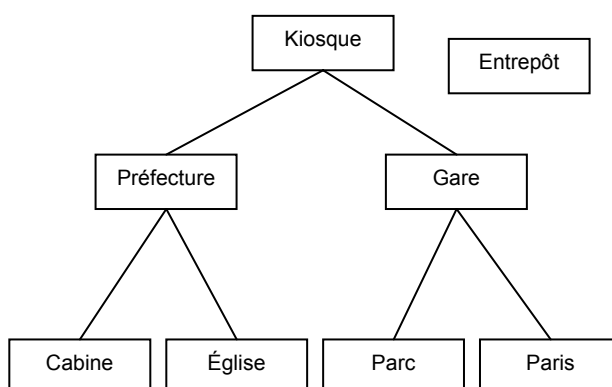
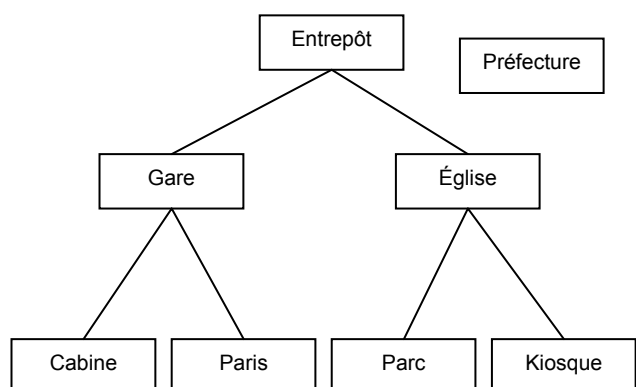
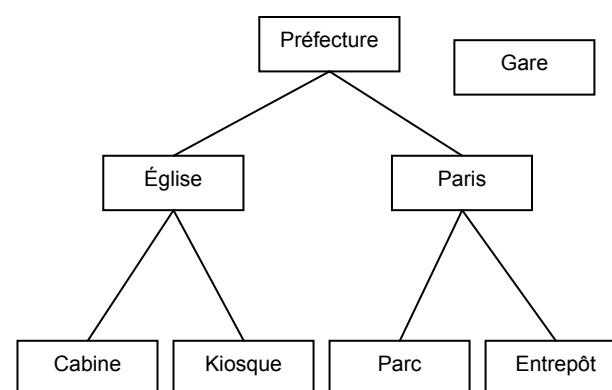
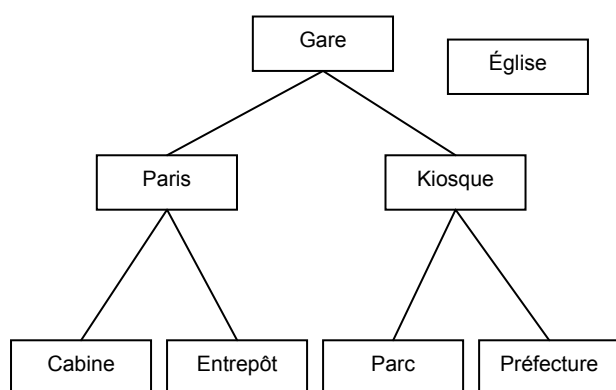
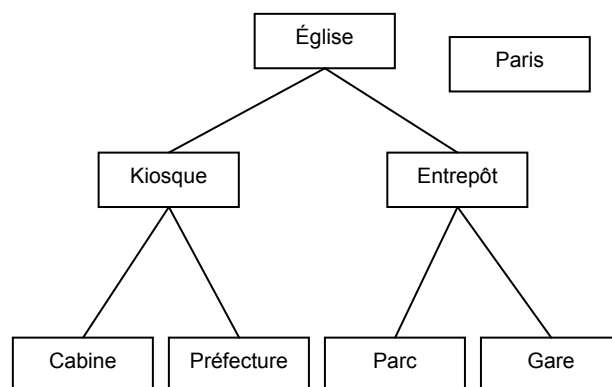
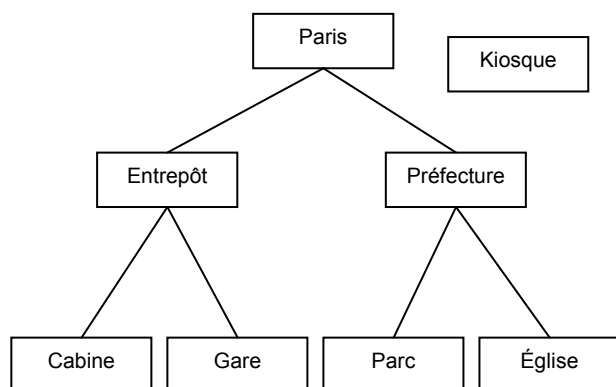
$$\Pr (DP / T, SBc) = R5 * T * SBc * (1P) * (1-SBc) * s * (1-g)$$

$$\Pr (DP / T, 1 - SBc) = R5 * T * (1-SBc) * (1P) * (1-SBc) * s * (1-g)$$

$$\Pr (DP / 1 - T) = R5 * (1-T) * (1P) * (1-SBc) * s * (1-g)$$

$$\Pr (ND) = R5 * T * SBc * (1P) * (1-SBc) * (1-s) + R5 * T * (1-SBc) * (1P) * (1-SBc) * (1-s) + R5 * (1-T) * (1P) * (1-SBc) * (1-s) + (1-R5)$$

ANNEXE 7 : ARBRES DES TÂCHES UTILISÉS DANS L'EXPÉRIENCE 6



ANNEXE 8 : PLAN EN CARRÉ LATIN DE L'EXPÉRIENCE 6

	GROUPES					
	G1	G 2	G3	G 4	G 5	G 6
Point de vue : IDENTITÉ						
Identité (église)	$n > 3$	n_{final}	NP	n	$n + 1$	$n - 1$
Localisation (kiosque)	NP	n	$n + 1$	$n - 1$	$n > 3$	n_{final}
Disparition (Paris)	n_{final}	NP	n	$n + 1$	$n - 1$	$n > 3$
Point de vue : LOCALISATION						
Identité (entrepôt)	n	$n + 1$	$n - 1$	$n > 3$	n_{final}	NP
Localisation (gare)	$n - 1$	$n > 3$	n_{final}	NP	n	$n + 1$
Disparition (préfecture)	$n + 1$	$n - 1$	$n > 3$	n_{final}	NP	n

ANNEXE 9 : NOMBRE DE RÉPONSES EN FONCTION DU RANG DU CHANGEMENT ET
DE LA COHÉRENCE POINT DE VUE / CHANGEMENT IMPLÉMENTÉ DANS LE
MODÈLE MPT (EXPÉRIENCE 6)

Pertinence du changement par rapport à la tâche				
	P		NP	
	Pertinence par rapport au sous-but en cours			
	oui	non		total
Cohérent				
DC	56	33	23	112
DP	16	19	6	41
ND	60	80	103	243
Incohérent				
DC	44	27	24	95
DP	17	19	7	43
ND	71	86	101	258

Note.: Étant donné que nous ne pouvons pas évaluer quelle est la part des non-détections (ND₁, ND₂, et ND₃) recueillies qui est due à un défaut de représentation de l'information (1 – R) ou à un défaut de « visual sensing » (1 – s), nous avons implémenté dans le modèle le nombre total de non-détection, et laissons au modèle le soin d'évaluer la part des non-détections qui découlent d'un défaut de représentation et d'un défaut de « visual sensing ».

ANNEXE 10 : ÉQUATIONS GÉNÉRÉES PAR LE MODÈLE MPT (EXPÉRIENCE 6)

Ces équations définissent la probabilité d'une catégorie de réponse (DC = détection correcte ; DP = détection partielle ; ND = non-détection) en fonction du type d'item (T = pertinent par rapport à la tâche ; 1 - T = non pertinent par rapport à la tâche ; SBc = pertinent par rapport au sous-but en cours de résolution ; 1 - SBc = non pertinent par rapport au sous-but en cours de résolution). Les données observées pour chaque catégorie de réponses figurent dans le tableau de l'annexe 9.

Expérience 6 – Changement consistant avec le point de vue (c)

$$\Pr (DC / T, SBc) = R * T * SBc * (1P) * PCc + R * T * SBc * (1P) * (1-PCc) * s * g + R * T * SBc * (1-1P)$$

$$\Pr (DC / T, 1 - SBc) = R * T * (1-SBc) * (1P) * PC + R * T * (1-SBc) * (1P) * (1-PC) * s * g + R * T * (1-SBc) * (1-1P)$$

$$\Pr (DC / 1 - T) = R * (1-T) * (1P) * PC + R * (1-T) * (1P) * (1-PC) * s * g + R * (1-T) * (1-1P)$$

$$\Pr (DP / T, SBc) = R * T * SBc * (1P) * (1-PCc) * s * (1-g)$$

$$\Pr (DP / T, 1 - SBc) = R * T * (1-SBc) * (1P) * (1-PC) * s * (1-g)$$

$$\Pr (DP / 1 - T) = R * (1-T) * (1P) * (1-PC) * s * (1-g)$$

$$\Pr (ND) = R * T * SBc * (1P) * (1-PCc) * (1-s) + R * T * (1-SBc) * (1P) * (1-PC) * (1-s) + R * (1-T) * (1P) * (1-PC) * (1-s) + (1-R)$$

Expérience 6 – Changement inconsistant avec le point de vue (i)

$$\Pr (DC / T, SBc) = R * T * SBc * (1P) * PCi + R * T * SBc * (1P) * (1-PCi) * s * g + R * T * SBc * (1-1P)$$

$$\Pr (DC / T, 1 - SBc) = R * T * (1-SBc) * (1P) * PC + R * T * (1-SBc) * (1P) * (1-PC) * s * g + R * T * (1-SBc) * (1-1P)$$

$$\Pr (DC / 1 - T) = R * (1-T) * (1P) * PC + R * (1-T) * (1P) * (1-PC) * s * g + R * (1-T) * (1-1P)$$

$$\Pr (DP / T, SBc) = R * T * SBc * (1P) * (1-PCi) * s * (1-g)$$

$$\Pr (DP / T, 1 - SBc) = R * T * (1-SBc) * (1P) * (1-PC) * s * (1-g)$$

$$\Pr (DP / 1 - T) = R * (1-T) * (1P) * (1-PC) * s * (1-g)$$

$$\Pr (ND) = R * T * SBc * (1P) * (1-PCi) * (1-s) + R * T * (1-SBc) * (1P) * (1-PC) * (1-s) + R * (1-T) * (1P) * (1-PC) * (1-s) + (1-R)$$

ANNEXE 11 : UN MODÈLE MULTINOMIAL DES NIVEAUX DE REPRÉSENTATION

DES SCÈNES VISUELLES COMPLEXES

A l'issue du chapitre 5, nous avons émis l'hypothèse de deux niveaux de représentation des informations visuelles de la scène, un niveau de représentation sémantique qui code l'intérêt sémantique des objets de la scène, et un niveau fonctionnel qui code la pertinence des objets de la scène par rapport à la tâche, indépendamment de leur intérêt sémantique.

Cette hypothèse peut être représentée en un modèle multinomial, illustré dans la figure A, ci-dessous.

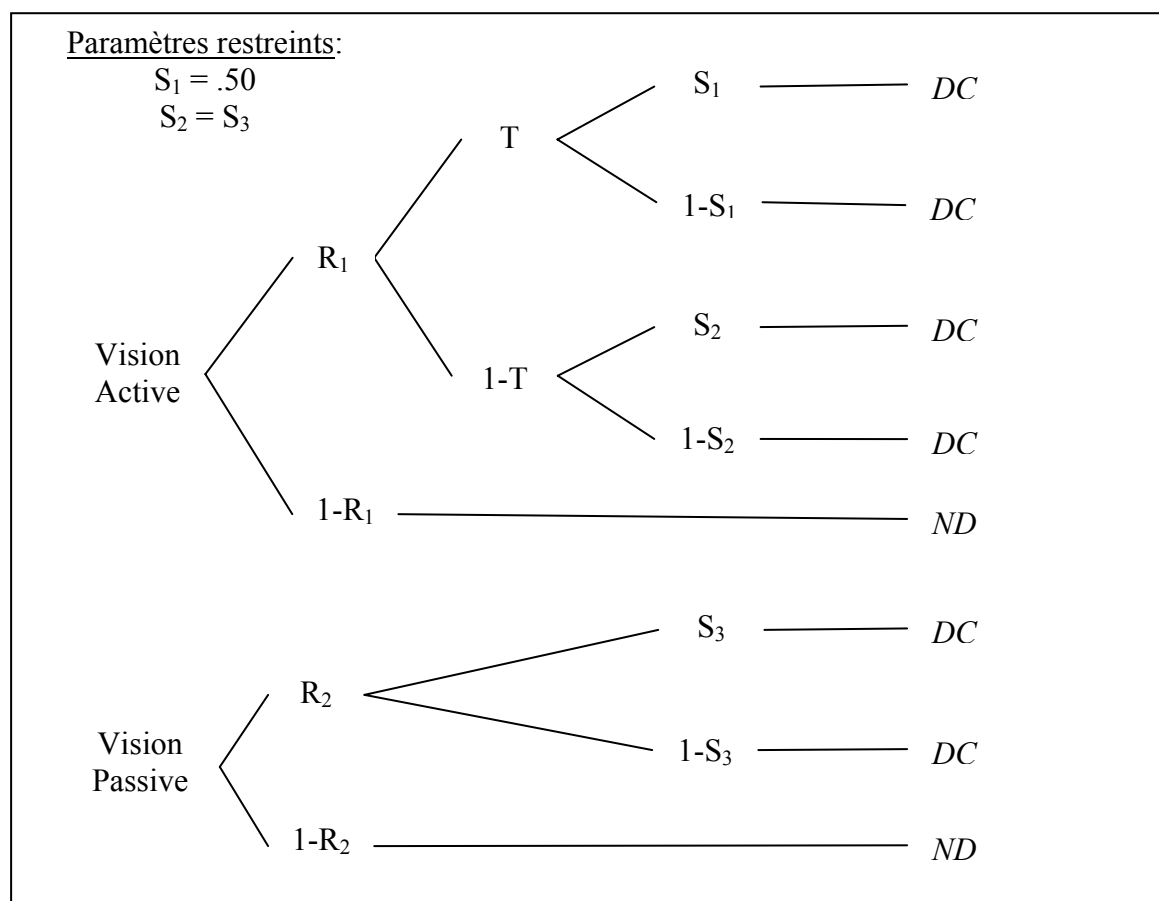


Figure A : Modèle multinomial des niveaux de représentation des scènes visuelles en mémoire. R_n = probabilité de représenter un item de la scène; T = probabilité de représenter un pertinent par rapport à la tâche ; S_n = probabilité de représenter un item d'intérêt central; DC = Détection correcte; ND = non-détection

Le modèle présenté dans la figure A est un modèle MPT joint, dans lequel le premier niveau reflète la condition expérimentale (vision active *vs.* vision passive) et le second niveau les réponses des participants (détection des changements *vs.* non détection des changements). Le modèle est conçu pour les tests de détection des changements avec deux réponses alternatives (détection *vs.* non détection) et quatre types d'items changés (pertinent par rapport à la tâche et d'intérêt central ; pertinent par rapport à la tâche et d'intérêt marginal ; non pertinent par rapport à la tâche et d'intérêt central ; non pertinent par rapport à la tâche et d'intérêt marginal). Comme le montre la figure A, le modèle propose deux conditions de vision (vision active et vision passive), et postule que les items de la scène peuvent être représentés en mémoire (R_n) ou non ($1 - R_n$). Le modèle postule également que les items peuvent être représentés selon leur intérêt sémantique (S_n) ou leur pertinence par rapport à la tâche (T).

En condition de vision passive (modèle du bas), le modèle considère que les items peuvent être représentés en mémoire (R_2) ou non ($1 - R_2$). S'ils sont représentés en mémoire, le modèle considère que les items de la scène peuvent seulement être représentés en fonction de leur intérêt sémantique¹¹³. Aussi, dans le modèle le paramètre S_3 correspond à la probabilité pour les items d'intérêt central d'être représentés (au sein d'un niveau de représentation sémantique), et le paramètre ($1 - S_3$) correspond à la probabilité pour les items d'intérêt marginal d'y être représentés. Si les items sont représentés en mémoire, le modèle prédit que leur changement devrait être détecté (DC), s'ils ne le sont pas, le modèle prédit une non-détection (ND).

En condition de vision active (modèle du haut), le modèle considère que les items peuvent être représentés en mémoire (R_1) ou non ($1 - R_1$). Par ailleurs, le modèle considère

¹¹³ En effet, en condition de vision passive, il n'y a pas de tâche à réaliser sur les scènes.

que deux niveaux de représentations existent, un niveau fonctionnel qui code la pertinence des items par rapport à la tâche, et niveau sémantique qui code leur intérêt sémantique. De fait, il propose deux paramètres distincts pour le traitement de la pertinence par rapport à la tâche et pour le traitement de l'intérêt sémantique. Ainsi, le paramètre T correspond à la probabilité pour les items pertinents par rapport à la tâche d'être représentés en mémoire (avec $1 - T$ correspondant à la probabilité pour les items non pertinents par rapport à la tâche d'être représentés en mémoire). Les paramètres S_2 et S_3 correspondent à la probabilité pour les items d'intérêt central d'être représentés en mémoire (avec les paramètres $1 - S_2$ et $1 - S_3$ correspondant à la probabilité pour les items d'intérêt marginal d'être représentés en mémoire). Enfin, le modèle considère (i) que les items pertinents par rapport à la tâche sont codés au niveau fonctionnel de représentation quel que soit leur intérêt sémantique, ce qui se traduit par : $S_1 = 1 - S_1$, et (ii) que lorsqu'ils sont non pertinents, les items sont codés à un niveau de représentation sémantique, qui code l'intérêt sémantique des items (central, marginal) de la même manière qu'en condition de vision passive, ce qui se traduit par : $S_2 = S_3$. Bien entendu, si les items sont représentés en mémoire, le modèle prédit que leur changement devrait être détecté (DC), sinon, il prédit une non-détection (ND).

1. Démonstration de l'identifiabilité globale du modèle

Afin que le modèle soit globalement identifiable, les paramètres doivent pouvoir être uniquement déterminés en fonction de la probabilité de la catégorie de réponse (Erdfelder, sous presse). Le tableau A donne les probabilités de réponse pour les différentes conditions et types de réponse. Chaque p_{ij} désigne la probabilité de réponse j à un item de type i . Le modèle

illustré ci-dessus a cinq paramètres indépendants, dont un fixé avec une probabilité constante (*):

$$\Omega 5 = [(R1, R2, T, S1^*, S)]$$

Les probabilités de réponse du tableau A peuvent être exprimées sous la forme de huit équations indépendantes, créées en sommant les probabilités de chaque branche de l'arbre illustré ci-dessus, pour chaque réponse :

$$p_{11} = R_1 T S_1 \quad (A1)$$

$$p_{21} = R_1 T (1 - S_1) \quad (A2)$$

$$p_{31} = R_1 (1 - T) S \quad (A3)$$

$$p_{41} = R_1 (1 - T) (1 - S) \quad (A4)$$

$$p_{12} = 1 - R_1 \quad (A5)$$

$$p_{51} = R_2 S \quad (A6)$$

$$p_{61} = R_2 (1 - S) \quad (A7)$$

$$p_{52} = 1 - R_2 \quad (A8)$$

A l'intérieur de chaque condition, la somme des probabilités des différentes catégories de réponses égale 1. Donc, une catégorie de réponse peut être exprimée en fonction des autres probabilités de réponses de cette condition. Par exemple, $p_{41} = 1 - p_{11} - p_{21} - p_{31} - p_{12}$. Nous choisissons d'exprimer les équations A4 et A7 en terme des autres probabilités de réponses, de leurs conditions (« Vision Active » et « Vision Passive ») respectives : cela restreint le nombre d'équations indépendantes à 6.

Il reste donc à résoudre les équations A1, A2, A3, A5, A6, et A8. Toutefois, ces équations peuvent être simplifiées et exprimées en fonction des autres probabilités de réponses, afin de réduire le nombre d'équations. Les équations A5 et A8 donnent respectivement:

$$R1 = 1 - p_{12} \quad (A9)$$

$$R2 = 1 - p_{52} \quad (A10)$$

L'équation A2 peut être réécrite et simplifiée comme suit:

$$p_{21} = R_1 T (1 - S_1) \quad (A2)$$

$$p_{21} = R_1 T - R_1 T S_1 \quad (A11)$$

Nous utilisons ensuite les équations A1 et A11 pour résoudre le paramètre T en terme de probabilité de catégorie de réponse. Les équations A1 et A11 peuvent être combinées pour résoudre le paramètre T

$$p_{11} + p_{21} = R_1 T S_1 + R_1 T - R_1 T S_1 \quad (A13)$$

L'équation A13 peut être, elle aussi, réécrite:

$$T = (p_{11} + p_{21}) / R_1 \quad (A14)$$

Les équations A1 et A11 peuvent être combinées afin d'exprimer le paramètre T en fonction des probabilités des catégories de réponses :

$$T = (p_{11} + p_{21}) / (1 - p_{12}) \quad (A15)$$

L'équation A3 peut être également simplifiée:

$$p_{31} = R_1 (1 - T) S \quad (A3)$$

$$p_{31} = R_1 S - R_1 S T \quad (A16)$$

Les équations A6 et A16 peuvent être combinées afin de résoudre le paramètre S

$$p_{31} + p_{51} = R_1 S - R_1 S T + R_2 S \quad (A17)$$

L'équation A17 peut être réécrite comme suit:

$$p_{31} + p_{51} = S (R_1 - R_1 T + R_2) \quad (A18)$$

R_1 , R_2 et T peuvent être remplacés dans l'équation A18 avec les équations A9, A10, et A15

$$p_{31} + p_{51} = S [(1 - p_{12}) - ((1 - p_{12}) (p_{11} + p_{21})) / (1 - p_{12}) + (1 - p_{52})] \quad (A19)$$

L'équation A19 peut être simplifiée

$$p_{31} + p_{51} = S [2 - p_{12} - p_{11} - p_{21} - p_{52}] \quad (A20)$$

Le paramètre S peut être isolé

$$S = (p_{31} + p_{51}) / (2 - p_{12} - p_{11} - p_{21} - p_{52})$$

Les paramètres R_1 , R_2 , T , et S , non maintenus constants, peuvent donc être exprimés en termes de probabilités des catégories de réponse. Le modèle est donc globalement identifiable.

Tableau A. Illustration des probabilités des catégories de réponses

Condition et type d'item		Type de réponse	
		Détection du changement (DC)	Non-détection du changement (ND)
Vision Active			
Pertinent, Central	p_{11}		
Pertinent, Marginal	p_{21}		p_{12}
Non Pertinent, Central	p_{31}		
Non Pertinent, Marginal	p_{41}		
Vision Passive			
Central	P_{51}		P_{52}
Marginal	P_{61}		

2. Le modèle est en faveur de deux niveaux de représentations des scènes visuelles

Le modèle illustré en figure A a été testé avec les données de l'expérience 2 (voir tableau B ci-dessous).

Tableau B : Fréquences de réponse en fonction de la condition et du type d'item. Les fréquences de non-détections implémentées dans le modèle sont soulignées (les valeurs entre parenthèses sont indicatives)

Condition et type d'item	Type de réponse	
	Détection du changement (DC)	Non-détection du changement (ND)
Vision Active		
Pertinent, Central	137	(151)
Pertinent, Marginal	127	(161)
Non Pertinent, Central	61	(227)
Non Pertinent, Marginal	32	(256)
Vision Passive		
Central	156	(420)
Marginal	91	(485)

Avec $N = 2,304$ (48 participants x 48 essais), 2 degrés de liberté et un risque alpha égal à .05, le test khi-deux d'ajustement indique une valeur $X^2(2) = .55$. Cette valeur est inférieure à la valeur critique donnée par la table du khi-deux $VC = 5.99$. Ce résultat indique que le modèle est en adéquation avec les données de l'expérience 2. Par contre, lorsque l'on contraint le modèle à considérer que la pertinence des items par rapport à la tâche ou leur intérêt sémantique ne sont pas des facteurs de la représentation en mémoire (*ie.*, quand on pose respectivement : $T = .50$ et $S_1 = S_2 = S_3 = .50$), le modèle n'est plus en mesure de rendre compte des données (voir tableau C pour les valeurs statistiques obtenues pour chacune de ces contraintes). Enfin, lorsque l'on contraint le modèle à considérer que la probabilité de représenter les items d'intérêt central ne diffère pas que les items soient par ailleurs pertinents par rapport à la tâche ou non (*ie.*, quand on pose $S_1 = S_2 = S_3$), le modèle n'est, là non plus, plus en mesure de rendre compte des données (voir tableau C).

Tableau C : test de l'adéquation du modèle aux données, en fonctions des contraintes apportées. X^2 = valeur obtenue au test de khi-deux ; dl = degrés de liberté ; VC = valeur critique donnée par la table du khi-deux

Contraintes	X ²	dl	VC	Modèle en adéquation avec les données?
Modèle initial				
S2 = .50	.55	2	5.99	OUI
S1 = S3				
Modèles alternatifs				
T = .50	81.92	3	7.81	NON
S1=S2=S3 = .50	26.88	3	7.81	NON
S1= S2= S3	8.88	2	5.99	NON

Ainsi, le seul modèle a pouvoir rendre compte de nos données est donc le modèle initial, illustré en figure A, qui postule deux niveaux de représentations des items contenus dans une scène visuelle naturelle, un niveau sémantique codant l'intérêt sémantique des items, et un niveau fonctionnel, qui code la pertinence des items par rapport à la tâche, indépendamment de leur intérêt sémantique.